TRANSPORTATION RESEARCH
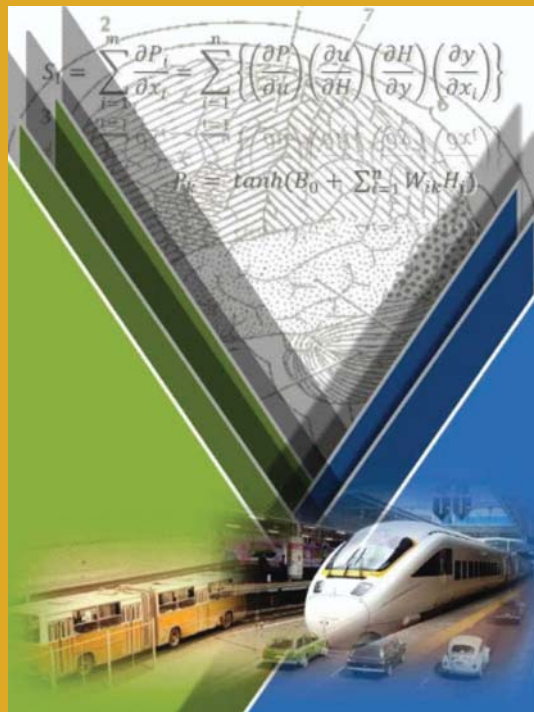# CIRCULAR

Number E-C168   **November 2012**

Artificial Intelligence Applications to Critical Transportation Issues

$$S_i = \sum_{j=1}^{m} \frac{\partial P_i}{\partial x_i} = \sum_{j=1}^{n} \left\{ \left(\frac{\partial P}{\partial u}\right)\left(\frac{\partial u}{\partial H}\right)\left(\frac{\partial H}{\partial y}\right)\left(\frac{\partial y}{\partial x_i}\right) \right\}$$

$$P_k = tanh(B_0 + \sum_{j=1}^{n} W_{jk} H_j)$$

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

Artificial Intelligence Applications to Critical Transportation Issues

Artificial Intelligence and Advanced
Computing Committee
Transportation Research Board
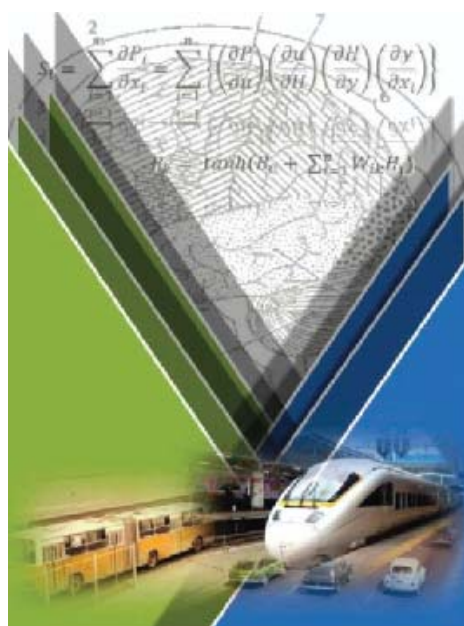
November 2012

Transportation Research Board
500 Fifth Street, NW
Washington, DC 20001
www.TRB.org

TRANSPORTATION RESEARCH CIRCULAR E-C168

The **Transportation Research Board** is one of six major divisions of the National Research Council, which serves as an independent advisor to the federal government and others on scientific and technical questions of national importance. The National Research Council is jointly administered by the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal.

The **Transportation Research Board** is distributing this Circular to make the information contained herein available for use by individual practitioners in state and local transportation agencies, researchers in academic institutions, and other members of the transportation research community. The information in this circular was taken directly from the submission of the authors. This document is not a report of the National Research Council or the National Academy of Sciences.

Foreword

As part of its mission, the Transportation Research Board's Artificial Intelligence and Advanced Computing Committee (ABJ70) serves as a technical forum on the application of artificial intelligence (AI) to transportation problems and disseminates information about AI applications that is credible and potentially useful to the transportation community. This is the second circular published by the committee. The first, *Circular E-C113: Artificial Intelligence in Transportation: Information for Application*, was published in 2007.

This circular, authored by members of the committee, encompasses three parts. The first part describes AI applications in transportation, including the advantages and limitations of AI. The second part encompasses 12 articles in five chapters describing five general AI areas: traffic operations, travel demand modeling, transportation safety and security, public transportation, and infrastructure design and construction. The third part presents a few of the committee members' thoughts on the future of AI research in transportation applications.

This circular is intended as an informational resource for transportation practitioners and managers about AI tools within these general areas. Each article details the types of problems to which the AI paradigm is best suited, its strengths and weaknesses, example applications, and guidelines for its application.

—Shinya Kikuchi
Chair, TRB Artificial Intelligence
and Advanced Computing Committee

PUBLISHER'S NOTE

This circular celebrates the contributions to TRB by **Shinya Kikuchi**, Charles E. Via Jr. Professor of Civil and Environmental Engineering and Director, Department of Civil and Environmental Engineering, Virginia Tech National Capital Region, who after six years is completing his term as Chair of the Artificial Intelligence and Advanced Computing Applications Committee.* This circular is the second in the series on artificial intelligence (AI) applications in transportation prepared under Prof. Kikuchi's leadership.

During his chairmanship, he has spearheaded the dissemination of AI knowledge through these circulars as well as through regular workshops at TRB Annual Meetings. Prof. Kikuchi has served on nine TRB committees, always making his mark by empowering his committee members to act and to push new ideas out to the TRB community and beyond. He has consistently drawn a high level of participation in TRB activities from researchers and practitioners around the globe.

* For additional information, see www.cee.vt.edu/people/faculty/TISE/shinya_kikuchi.html.

Acknowledgments

Contents

Application Area 4: Public Transportation

Application Area 5: Infrastructure Design and Construction

**THOUGHTS ON THE FUTURE OF ARTIFICIAL INTELLIGENCE
AND TRANSPORTATION**

Why Artificial Intelligence?

Difference Between Artificial Intelligence and Traditional Methods

Henk van Zuylen
Delft University of Technology

The domain of artificial intelligence (AI) is wide. Originally AI was defined as the discipline of computers that show intelligent human behavior. Now AI also refers to computer systems that show complex behavior similar to living systems like swarms, ant colonies, microbiology, and neural systems. There is a distinction between "'strong" AI—computer functions that really have strong similarities with intelligent human reasoning and show some kind of self awareness—and "weak" AI—computer applications that deal with limited application areas and contain some practical knowledge and seem to have some intelligent features, such as expert systems and heuristic search algorithms. The focus of engineering applications of AI such as the ones considered in this circular is more on weak AI. However, it should be noted that the boundaries between strong and weak AI are not sharp and often AI systems move from one type to the other.

A general distinction between "ordinary" computer systems and AI is the complexity of the AI computer systems, including

- The complexity of AI computer programs themselves,
- The complexity of the performance of the algorithms, and
- The complexity of the subject of the AI system—that is, the data that have to be processed, the part of the reality that has to be modeled, or the behavior that has to be shown.

However, non-AI programs can also be quite complex, e.g., a computer model of the flow of air around an airplane is very complicated; however, it is based on the Navier–Stokes equations, well-known physical laws written in mathematical formulas. Therefore, it is transparent how the dynamics of the flow are determined, notwithstanding the complexity of the shape of the boundary surfaces around which the air flows and the turbulence of the flow. Such a program is not considered AI, because it is based on a well-established analytical theory.

Some applications are considered as AI because they use an analogy from biology. Such is the case with neural networks. The difference between traditional regression methods and neural networks is not significant. Both methods describe the relationship between independent and dependent variables in a more or less heuristic way. A causal model is not necessary. A neural network can learn this associative relation from observed data in a manner similar to parameter calibration in traditional regression models. The structure of a neural network bears some similarity with the structure of neurons in living creatures, and the way neural networks are trained bears some similarities with learning. An AI system can produce very complex, pseudo-intelligent behavior without being very complex itself. The famous program ELIZA, made by Joseph Weizenbaum, has the amazing function of creating a dialogue with the user similar to the dialog of an intelligent psychotherapist. If you read the program you see a clear structure of operation. The program mimics a Rogerian therapist by asking questions and giving answers based on the input entered by the user. It does it in such a way that the user has the impression

that the computer understands his questions and answers and gives a human-like intelligent reaction. For somebody who knows the structure of the program, the dialogue from the side of the computer is predictable because the way a new question or answer is produced is determined by only 1,081 lines of code. The program certainly can be seen as a kind of strong AI because it is adaptive and processes verbal information in a way humans do.

An important class of AI is formed by programs that can process and produce language. Procedural programming languages process numbers and logicals. For processing language the procedures are much more complicated than basic mathematical procedures. The logical reasoning with language elements is the basis for fuzzy reasoning, which involves operations on concepts that do not have a sharp definition or belonging to a set, but instead a certain level of membership with one or more collections of more or less similar concepts.

Programs that can process visual data can be considered as AI, especially when these programs are able to recognize and deal with shapes. Shape recognition and manipulation are typical forms of intelligent behavior. Computer applications that process graphical data in order to extract special features like faces, handwriting, number plates, vehicle type, etc., have been widely used. The complexity lies in the analysis of the image data and the identification of relevant characteristics.

Robotics is another AI application that is becoming an increasingly common computer application. Robots that operate in a controlled environment do not require much in the way of system intelligence. However, robots that operate in an unspecified environment with complex, uncontrolled conditions must adapt their behavior to their environment and therefore need features to recognize the structure of the environment and then develop operational plans. These kinds of robots clearly behave in an intelligent way.

Another mechanism by which AI programs exhibit complex and intelligent behavior arises from the cooperative features of simple procedures. The cooperation leads to emergent intelligent behavior. A familiar example is the behavior of ants in a colony. Each worker ant in a colony has a very simple task, i.e., searching for food. When it finds food, the ant goes back to the anthill, marking the path as it returns. This marked path stimulates other ants to forage in the same direction. This process has been mimicked in search algorithms to look for optimal solutions to complicated systems. Such intelligent search techniques have been employed to provide efficient search programs. However, not all cooperative systems yield emergent intelligent behavior, and the intelligence of the AI program designer is necessary to create an efficient, intelligent search program. The fact that the behavior of such cooperative, self-organizing systems is nearly unpredictable is an important consideration in their applications.

Genetic algorithms are another class of AI search commonly used for complex systems that cannot easily be optimized by analytical search methods. Similar to ant colony-based techniques, genetic algorithms have a complex, nondeterministic characteristic. This characteristic result from the paring of an opportunistic search direction coupled with a semi-stochastic component for the development of candidate solutions. However, the question whether a global optimum has been found by such a heuristic, stochastic search process cannot be answered. Nonetheless, in practice this search procedure from AI has proven to be simple, fast, and effective. Little special knowledge is needed to apply genetic algorithms, and users apply the technique to problems where a proof of global optimality is not essential.

One criticism, however, that is sometimes leveled against heuristic search algorithms such as genetic algorithms stems from their stochastic and uncontrollable nature. Many years ago I was the student of Edgar Dijkstra, who was a pioneer in computer science and taught the course

"The Art of Programming." He hammered on his students the need of well-structured algorithms with a clear structure, invariants, logical preconditions, terminating conditions, and correctness proof, and I sometimes wonder whether Professor Dijkstra would have approved many AI applications, such as swarm intelligence, ant colonies, and genetic algorithms. Some would thus argue that optimization algorithms from AI that are uncontrollable and heuristic might be a fast way to solve a complex problem, but sometimes it is also the lazy way. They would thus proceed to argue that a better analysis may sometimes lead to a better optimization method than the "easy" AI-based method. It is our perspective, however, that it may be often advantageous to take advantage of both AI methods as well as the insights that are gained by traditional analytical solution methods; by combining AI and analytical methods, powerful problem-solving methodologies may be developed.

To summarize: AI are characterized as computer applications that show features similar to intelligence of living creatures, such as evolution, collaborative behavior, processing of natural language, dealing with spatial structures, logical reasoning with fuzzy concepts, and learning. Especially during the initial development phase the specific AI character remains present. When the application becomes more established it often becomes a common computer application and is integrated in larger systems. The typical intelligent features often become a part of the functionality of the larger system.

The fact that AI programs in the past were programmed in a declarative form while non-AI programs were written in a procedural way is no longer relevant since the modern programming environments have integrated both features.

The following articles in this circular will elaborate on the advantages and limitations of AI when applied to transportation.

Advantages and Limitations of Artificial Intelligence

MASHRUR CHOWDHURY
Clemson University

ADEL W. SADEK
The State University of New York at Buffalo

Artificial intelligence (AI) applications are utilized to simulate human intelligence for either solving a problem or making a decision. AI provides the advantages of permanency, reliability, and cost-effectiveness while also addressing uncertainty and speed in either solving a problem or reaching a decision. AI has been applied in such diverse realms as engineering, economics, linguistics, law, manufacturing and medicine, and for a variety of modeling, prediction, and decision support and control applications (*1*). One of the most promising applications of AI has been its rigorous use in the Internet such as in search engines (*2*). Although the efficacies of AI are significant, as with any application they are limited in both capability and functionality. These limitations will be presented later in this article. Before describing AI's limitations, this article will briefly survey some of AI's advantages.

In an organization in which human intelligence is tied to a particular person or a group of people, AI applications can provide permanency that prevents the knowledge from being lost when the individual or the group members retire or are no longer available to the organization. The life of the knowledge encapsulated in an AI framework could be as long as the relevance of the problems and decision scenarios remain unchanged. AI also enables the development of a learning capability which can be utilized to further prolong the life and relevance of the application. Learning from real-world success and failure is an enabling feature of AI tools known as "reinforcement learning" and is advantageous in that it increases the reliability of the tools with their increased use in applications (*2*).

The broad application of any tool only occurs when its reliability has been established, and AI has already proven to be quite reliable in many different applications because of its ability to simulate human intelligence in a reasoning process. Like many automations, AI supports cost minimization as it enables reduction on the need of personnel time. An agency can reduce significant staff time by adopting appropriate AI applications in the decision-making process, thus reducing operational costs.

As decisions must often be made under obvious uncertainties (i.e., with incomplete and uncertain knowledge), AI methods are suitable when a direct mathematical relationship cannot be established between cause and effect. AI models capture the uncertainty between real-life cause and effect scenarios by incorporating available knowledge with probabilities and probability inference computations (*3*). AI methods are also capable of dealing with both qualitative as well as quantitative data, a feature that most strictly analytical methods lack.

Depending upon the computational time in terms of algorithmic complexity and processor capacity, AI tools can facilitate faster decision making by automating the decision-making process. Through data gathering and screening, processing, and decision making, AI can support faster solutions to complex problems.

In transportation, numerous research and applications have demonstrated many of the above advantages of AI in general, and significant research has provided evidence of the advantages of specific AI tools. Some examples of AI technology currently in use include converting traffic sensors into intelligent agents that can automatically detect and report traffic accidents or predict traffic conditions (*4*). More recently, researchers have found AI to be more reliable in assessing and predicting traffic conditions, based on microscopic traffic data collected from vehicles on their path, as envisioned in the vehicle–infrastructure integration or the connected vehicle program, compared to many other existing algorithms. Utilizing microscopic traffic data, transportation security is another realm in which AI can be of significant use (*5*). Here, AI tools can be applied to the identification of security breaches, and in the development and management of automated response and control plans.

The well-known advantages and efficacies of AI make them particularly useful in the development and management of transportation systems. Specifically, in intelligent transportation systems, real-time sensing, detection, response, and control are of paramount importance, and AI can be utilized effectively in all of these applications. Indeed, a distributed traffic sensor and control network is perhaps the future of real-time traffic management and control. Here, distributed sensor networks, comprised of different level of intelligent sensor networks, automatically detect and respond to incidents and control roadway network as needed. Such an intelligent sensor network comprised of AI tools can support the development of the next generation traffic management system. With the many advantages of AI tools, we expect their wider adoption in different areas of transportation.

Nevertheless, one should not forget that, like any other tool, AI methods have their limitations. One major criticism of many AI paradigms (e.g., neural networks), which was previously alluded to in the article by van Zuylen, is that they are often regarded as black boxes that merely attempt to map a relationship between output and input variables based on a training data set. This also immediately raises some concerns regarding the ability of the tool to generalize to situations that were not well represented in the data set. One solution that has been proposed to address the black box problem is the combination or integration of multiple AI paradigms into a hybrid solution (e.g., combining neural networks and fuzzy sets into neuro-fuzzy systems) or coupling AI paradigms with more traditional solution techniques.

Another limitation of AI-based search methods, such as genetic algorithms and ant colony optimization, is that they are never guaranteed to reach the "optimal" solution. Also when using AI-based search methods to solve a problem, it is often hard to gain true insight into the problem and the nature of the solution, as is possible for example when using mathematical programming methods. The inability to quickly do sensitivity analyses is an important example of this limitation. The counter argument for the inability to guarantee optimality is that for hard optimization problems that defy solution using traditional optimization and mathematical programming methods, a "solution" is still better than "no solution" at all. Moreover, there is significant empirical evidence to suggest that AI-based search methods do yield "good" solutions in most cases. For gaining insight into the problem, the model may have to be rerun multiple times to assess the sensitivity of the solution to the various assumptions and parameters of the problem, which may be somewhat demanding from a computational resources or runtime standpoint.

A third limitation associated with the use of AI methods to solve a given problem stems from the fact that, for several AI methods, there is currently little guidance on how to decide upon the best values to use for a given method's tuning parameters. Considering neural

networks, for example, the analyst needs to make some very important design decisions including the topology of the network to use, the number of hidden layers, the number of neurons to use in the hidden layer, the type of trigger functions to use in the neurons, among others, before applying this method to a given problem. In current practice, an analyst often has to go through a tedious trial-and-error procedure to select appropriate values for these parameters. The same is true for genetic algorithms, where the analyst needs to decide upon the population size, the number of generations, and the algorithm's other control parameters such as the probabilities for applying the mutation and crossover operators.

With respect to automation in general, and AI is no exception, there is also another limitation or a challenge related to the issue of potential liability. For example, if in the future, AI methods are used to build partial or full autonomous vehicles, for example, and if an autonomous vehicle is involved in a crash, who should be held liable in such a case? While this limitation is not technical in nature, it is a serious issue that needs to be addressed.

Finally, if we are to specifically focus on AI applications in transportation, one has to note that there is currently still a lot of skepticism among transportation practitioners regarding the ability of AI to help solve some of the problems they face. One reason behind such skepticism may in fact be attributed to the very name of the field itself—"artificial intelligence"—and the rather unrealistic goals for the field adopted by early AI researchers. As this committee has pointed out in *Transportation Research Circular E-C113: Artificial Intelligence in Transportation: Information for Application*, our purpose, as transportation researchers and professionals, is not to develop a general-purpose problem solver, but rather to address real transportation problems that have defied solution using classical solution methods.

The following part of this circular will therefore briefly survey the wide range of transportation problems and applications areas where AI techniques have been applied. The part will also identify additional opportunities for applying AI methods. Specifically, five broad application areas where AI has been applied will be reviewed. These are (*a*) traffic operations; (*b*) travel demand modeling; (*c*) transportation safety and security; (*d*) public transportation; and (*c*) infrastructure design and construction.

REFERENCES

1. Mellit, A., and S. Kalogirou. Artificial Intelligence Techniques for Photovoltaic Applications: A Review. *Progress in Energy and Combustion Science,* Vol. 34, 2008, pp. 574–632.
2. Russell, S., and P. Norvig. *Artificial Intelligence: A Modern Approach*, 2nd edition. Pearson Education, Inc., 2003.
3. Patterson, D. *Introduction to Artificial Intelligence and Expert Systems.* Prentice Hall, Inc., 1990.
4. Schleiffer, R. Intelligent Agents in Traffic and Transportation. *Transportation Research Part C,* Vol. 10, 2002, pp. 325–329.
5. Fries, R., M. Chowdhury, and J. Brummond. *Transportation Infrastructure Security Utilizing Intelligent Transportation Systems*. John Wiley & Sons, 2008.

Artificial Intelligence and Key Transportation Applications Areas

Traffic Signal Timing and Optimization

Yunlong Zhang
Texas A&M University

Yuanchang Xie
University of Massachusetts, Lowell

Artificial intelligence (AI) methods have been frequently employed to solve complicated control and optimization problems in many different disciplines. Traffic signal control and optimization is no exception. A wide range of AI methods have been introduced into this area to develop innovative traffic control strategies and fine tune signal timing plans and parameters. These AI applications are summarized into the following four categories and detailed in the remaining of this article: evolutionary algorithms, fuzzy logic control, artificial neural networks control, and reinforcement learning and agent-based control.

EVOLUTIONARY ALGORITHMS

Finding the best traffic signal timing plans is often formulated as a mixed-integer linear programming or dynamic programming model. These models usually involve the simultaneous optimization of phasing sequence, split, cycle length, and offset, and it is often difficult to find the globally optimal solutions to such models within a reasonable amount of time using exact algorithms. AI methods, specifically evolutionary algorithms, have been extensively researched as an alternative to exact algorithms to address this issue. These evolutionary algorithms include hill climbing, genetic algorithms (GAs), particle swarm optimization (PSO), and ant colony optimization (ACO).

Hill climbing is a local search method for solving optimization problems. It starts with a randomly generated solution and utilizes an iterative process to improve it. At each iteration, the current best solution is selected to generate the next candidate solution (similar to offspring in GAs), which is kept only if it can result in a better objective function value. Hill climbing has been implemented in some commercial traffic signal timing programs such as PASSER V-03 and TRANSYT-7F. A major problem of hill climbing is that it may easily get stuck in some local optimal solutions. Several strategies have been proposed to improve its global search ability, such as introducing a tabu list and using random restarts (*1*). However, additional research is needed to assess the effectiveness of these strategies in improving traffic signal timing plans.

GAs have been very popular for traffic signal timing and optimization (*2–7*). A typical example is the research of Park et al. (*3*). They proposed a GA approach for traffic signal optimization under oversaturated conditions. The GA approach has two major components: a GA optimizer and a mesoscopic simulator. The GA optimizer generates an initial population consisting of a number of feasible solutions. Each solution (also called chromosome in GA) represents a traffic signal timing plan with cycle lengths, splits, offsets, and phasing sequence specified. The fitness values of these solutions are evaluated using the mesoscopic simulator.

Based on the fitness values, the GA optimizer performs selection, crossover, and mutation operations on the population and searches for the best signal timing plans. A key issue in applying GA for traffic signal optimization is how to code OR decode cycle lengths, splits, and offsets. A fraction-based coding scheme was used by Park et al. (*3*), which was successfully applied to a two-intersection network. If the traffic network size increases, it is possible that GA may need considerably more computation time and the optimality of the solution may also deteriorate. Such decreases in efficiency are partially attributed to the many additional constraints (e.g., minimum green) and the large search space. Similar to the hill climbing method, GA optimization is implemented in both PASSER-V 03 and TRANSYT-7F.

Recently, PSO (*8*) has also been adopted for traffic signal optimization (*9*). Similar to GA, the PSO generates a swarm (set) of initial solutions at the beginning. Each solution is referred to as a particle and has its own location. Instead of applying crossover and mutation operations only to the selected chromosomes as in GA, the PSO updates every particle's location at each iterative step. As the iterative process proceeds, each particle will have many different locations. The PSO keeps track of the best locations of each particle and the best location of all particles in the entire swarm. At each iterative step, these best locations, a randomly generated new location, and the current particle locations are used to update each particle's location. When the termination criterion is met, the best location of all particles in the swarm is output as the final solution. Wei et al. applied the PSO to fine tune the parameters of a fuzzy-logic traffic signal controller and found that it can effectively improve the performance of the original fuzzy-logic controller (*9*).

Another interesting evolutionary method for traffic signal timing is ACO, which was introduced by Putha et al. for oversaturated traffic signal coordination (*10*). In this study, the ACO was compared with GA based on a 20-intersection network without considering turning movements. Comparable results from the two methods were reported. The authors further pointed out that the structure of the ACO algorithm makes it particularly suitable for parallel computing, which can substantially shorten the computation time.

Evolutionary algorithms have promising potential for finding the globally optimal solutions to traffic signal timing and optimization problems. However, these algorithms have their respective limitations. It may take a considerable amount of time for them to find the globally optimal solution or a reasonably good solution, especially when the number of decision variables and constraints are large. Thus, the applications of evolutionary algorithms in traffic signal control have to date been primarily for small-scale or off-line optimization problems. With continued advancements in computational technology and new developments in evolutionary algorithms, it is expected that more applications will emerge that apply evolutionary algorithms for large-scale or real-time (*7*) traffic signal optimization and coordination in the future.

FUZZY LOGIC CONTROL

Numerous traffic signal timing studies have been conducted based on fuzzy logic (*11–16*). This makes it very challenging to perform an exhaustive review of existing publications. Thus, in this circular, we choose to focus on several representative papers and highlight some of the major findings and issues that may deserve further exploration. Most existing fuzzy logic traffic control studies use queue lengths and traffic arrivals as the input, and the control action usually is to

either extend or terminate the current green phase. The following is a sample fuzzy rule proposed by Zhang et al. (*14*):

> **if** current queue length is {long} **and** arrival rate is {high} **and** queue length for
> the next movement is {short}, **then** extension is {yes}

In four states, {short, medium, long, very long} are considered for current queue length and queue length for the next movement; three states, {low, medium, high} are considered for arrival rate; and in total, 48 fuzzy rules are defined (*14*). The control decision is made based on the states of the current and next movements. The phasing sequence is fixed. Obviously, for better control performance, it is desirable to consider variable phasing sequences and also include the states of other movements. However, including them will significantly increase the number of fuzzy rules to be specified, making the fuzzy rules considerably more complicated. This probably explains why many previous studies (*11–13*) either ignore turning movements or focus on modeling one-way streets. In this way, only two phases need to be considered for each intersection and defining fuzzy rules becomes much easier.

More realistic multiple-phase control (*14–16*) strategies are considered in recent studies. Trabia et al. developed a two-stage fuzzy logic control for a four-approach isolated intersection with both through and left-turn movements (*15*). The two-stage fuzzy logic controller is compared with an optimized actuated traffic controller based on average delay and percentage of stopped vehicles. In their research, four control phases are considered, including two left-turn phases and two through phases. The proposed two-stage fuzzy logic control is applied only to the through movements, while the regular actuated control is adopted for left-turn movements. The first stage is to calculate the traffic intensities of through movements based on maximum flow and queue length. The second stage determines whether the current green phase should be extended or terminated given the traffic intensity outputs. The authors discussed that the detector configurations of most real-world intersections make it difficult to measure left-turn queue lengths, which are required by the two-stage fuzzy logic controller. If queue length and flow information is also available for left-turn movements, they can also be controlled by the two-stage fuzzy logic method.

Zhang et al. also developed a fuzzy logic controller for isolated intersections with two-way streets and left-turn movements (*14*). They assumed that an upstream passage detector is installed in every lane so that accurate queue and flow data can be obtained and used as the controller input. As described previously, the proposed fuzzy logic controller only takes the current green phase's queue length and flow and the next green phase's queue information as the input. In both studies (*14, 15*), fuzzy logic is used to decide whether to extend the current green phase or not. The phasing sequence optimization is not explicitly taken into account.

Motivated by the importance of phasing sequence optimization, Murat and Gedizlioglu proposed an FLMuSiC fuzzy logic control consisting of two modules (*16*). The first module takes three input values: (*a*) the longest queue in red signal (LQRS); (*b*) arrivals to junction during green signal; and (*c*) green time indicator representing the ratio of remaining green time for the current phase. The outputs of the FLMuSiC are {more decrease, decrease, do not change, increase, more increase}, which are different from those used in the previous two models (*14, 15*). It appears that the FLMuSiC makes a one-time split adjustment when a phase starts receiving green signal. The second module of the FLMuSiC performs phasing sequence selection also based on three input values: (*a*) LQRS; (*b*) longest vehicle queue in next phase; and (*c*) red

time of the longest queue. There are only two outputs for the second module, which are change and do not change. The authors demonstrated how this strategy can be used for phasing sequence selection of a three-phase and single-ring control. However, no details were provided regarding how it can be used for dual-ring and more complicated controls (*16*). Murat and Gedizlioglu compared the proposed method with an actuated control. They reported that the FLMuSiC outperforms the actuated control for high traffic volume while underperforms it for low traffic volume (*16*).

Most research on fuzzy logic traffic signal control has been focused on isolated intersections. It appears that only Chiu and Chand considered fuzzy logic for network traffic signal control (*12*). They applied fuzzy logic control to a nine-intersection traffic network with one-way streets. Additional research is still needed to explore the potential of fuzzy logic for traffic signal coordination considering two-way traffic. Another promising research area is the integration of fuzzy logic and traditional control methods (*15*).

One advantage of using fuzzy logic for traffic signal control is that it needs minimal computational resources. Also, fuzzy logic makes it possible to incorporate ambiguous human experience into traffic signal control. Such experience might be difficult to describe by exact mathematic models. All these reasons explain the popularity of fuzzy logic in traffic signal control. Although many research papers have been published in this area, few real-world traffic control applications using fuzzy logic can be found (*17*). This probably is because practitioners prefer more straightforward methods such as pretimed and actuated control strategies. Configuring and testing pretimed and actuated controllers are relatively easy compared to specifying the many fuzzy logic rules and membership functions. Another reason may be that there are no generally agreed guidelines for defining fuzzy rules and membership functions. Different studies use quite different inputs and fuzzy rules as described previously. Additionally, most intersection detectors are laid out for actuated control. It would be very costly to rebuild the infrastructure for fuzzy logic control.

ARTIFICIAL NEURAL NETWORKS CONTROL

Compared to fuzzy logic, relatively few research has been conducted using neural networks for traffic signal control (*18–22*). In several cases, neural networks are integrated with other AI methods for traffic signal timing. Some of these integrated studies (*23, 24*) are reviewed in the next section.

Nakatsuji and Kaku developed a multilayer neural network model for self-organizing traffic control (*18*). The proposed model uses phase splits as the input and measures of effectiveness (MOE) such as queue lengths as the output. It is first trained by a back-propagation algorithm to establish an input–output relationship between the phase splits and MOEs. The trained model is then applied to a three-intersection network. A stepwise method consisting of a Cauchy machine and a steepest descent feedback method is employed to adjust the phase splits, in order to optimize (minimize) the squared sum of queue lengths on all inflow links. The authors concluded that the developed neural network control model generates approximately the same phase splits and queue lengths as some analytical tools, demonstrating the feasibility of using neural networks for traffic signal control.

Gilmore and Abe developed two separate neural network systems that were integrated into an advanced traffic management system (*19*). The first system is a Hopfield neural network

model for adaptive traffic signal control, and the second system is a back-propagation model for predicting traffic congestion. These two models are evaluated based on traffic simulation.

Hua and Faghri developed a neural signal control system (NSCS) for isolated intersection control based on multilayer neural network architecture (*20*). The authors concluded that the developed NSCS is able to better adapt to changes in traffic flow patterns than existing signal control methods. It can also provide more flexibility than traditional traffic signal controllers in optimizing different measures of effectiveness. The authors conducted a comprehensive evaluation of the NSCS and compared it with a GA-based signal operations analysis package (SOAP) based on computer simulation (*25*). The results show that the NSCS method outperforms the SOAP control for all MOEs considered.

Saito and Fan proposed a two-phase traffic signal control method (*21*). It consists of a level-of-service (LOS) evaluation tool based on artificial neural networks and an optimal traffic signal timing model (OTSTM). The OTSTM is a heuristics optimization model and is not based on neural networks. It is used to search for the best traffic signal control parameters. These parameters are then assessed by the LOS evaluation tool. This research is similar to the one conducted by Nakatsuji and Kaku, who also adopted neural networks for evaluating different control plans (parameters) (*18*). A major difference is that Nakatsuji and Kaku used neural networks directly to fine tune traffic control parameters (i.e., phase splits), while Saito and Fan adopted a heuristic model (i.e., OTSTM). For both studies, one thing in common is that they all employed neural networks to establish a relationship between control parameters and traffic control performance. Thus, the controller does not need to directly rely on any microscopic or mesoscopic traffic simulators to estimate the performance of a specific control plan. Although satisfying results were reported in both studies, it is important to give adequate attention to the potential generalization problems with neural networks. Given modern computers, microscopic or mesoscopic simulation models can generate reliable MOE estimates at very fast speeds. It would be interesting to find out whether it is worth trying to estimate them by neural networks.

REINFORCEMENT LEARNING AND AGENT-BASED CONTROL

Reinforcement learning and agent-based control recently have attracted considerable attention from transportation researchers. In reinforcement learning, the traffic signal controller is referred to as agent and the traffic to be controlled is referred to as environment. A system consists of a group of agents that work relatively independently and also interact with each other is called a multiagent system. Similar to neural networks, agents need to be trained by reinforcement learning algorithms before they can actually be used. Different from neural networks, the training of an agent is a dynamic process based on the continuous interactions between the agent and environment, not a fixed set of paired input–output training samples.

Reinforcement learning essentially is based on Markov decision process (MDP) theory. An MDP is a discrete time stochastic process characterized by a set of states (S), actions (A), reward function (r), and state–transition function (p). For intersection traffic signal control, the states can be defined by queue lengths and signal status; the actions are the available control strategies for each state; the reward function defines the immediate reward of each action under a specific state; and the state–transition function defines the probabilities for the system to shift from one state to another given the current state and action taken. To solve traffic signal control

problems formulated as MDPs, a key step is to find the optimal value function based on the following Bellman optimality equation.

$$V^*(s) = \max_{a \in A(s)} \left\{ \sum_{k \in S} p_{sk}^a r_{sk}^a + \gamma \sum_{k \in S} p_{sk}^a V^*(k) \right\}$$

where $a \in A(s)$; $s \in S$ is the current state and $k \in S$ represents the next state(s) after action a is taken; p_{sk}^a and r_{sk}^a are the transition probability and reward, respectively, from state s to state k after action a is taken; and $\gamma \in [0,1)$ is a discount factor. After $V^*(s)$ is found, the signal control problem simply becomes identifying the current system state s and applying the control action $a \in A(s)$ that leads to the optimal value function $V^*(s)$. The optimal value function can be found by a number of reinforcement learning algorithms, including Q-Learning and Actor–Critic Learning (*26*).

Thorpe conducted one of the pioneering studies on traffic signal control using reinforcement learning (*27*). In his study, a SARSA reinforcement learning algorithm is developed to train each intersection control agent and the learning result is stored in a Q-table. Each cell in the Q-table corresponds to an action value $V(s,a)$ for a state–action pair (s,a). Thorpe tested the SARSA control method on a 4x4 grid traffic network without left-turn movements. Each of the 16 intersections is controlled by one agent, and there is no coordination explicitly considered. One key issue for reinforcement learning applications is how to define states. In Thorpe's study, four different methods are used to define the states, which are (*a*) vehicle count representation; (*b*) fixed-distance representation; (*c*) variable–distance representation; and (*d*) count–duration representation. Another very important issue that affects signal control agents' performance is the definition of reward. Thorpe used two different definitions of reward. For the first definition, if at each decision point the environment state is not the goal state (all vehicles are cleared in the goal state), then the value for the action taken at the previous decision point is updated by –1. For the second definition, the reward is calculated as

$$r = \text{cons} + \text{moved} - \text{stopped}$$

where cons is a constant value that is set to –3 in Thorpe's study; moved is the number of vehicles that have passed the intersection from approaches being given green signal; and stopped is the number of vehicles that have been stopped due to a red signal in the last interval.

Abdulhai et al. proposed an adaptive traffic signal control strategy based on Q-Learning (*28*), which is a typical reinforcement learning method detailed in (*25*). In their study, Abdulhai et al. discussed how to apply Q-Learning to both isolated intersection and arterial traffic control, and provided testing results for an isolated intersection. The isolated intersection considered does not have turning vehicles and has a fixed cycle length. In each cycle there is one decision to make: either switch the signal or maintain its current status. The total delay accumulated between two consecutive phase switching points is used as the reward. The queue lengths on each approach and the elapsed time since last phase switching are used to define states. The authors proposed a cerebellar model articulation controller technique for storing and generalizing the learned value function. In their study, Abdulhai et al. also described a general framework for arterial and network traffic control using Q-Learning. They suggested including the queue

information from adjacent intersections to define the state of the current intersection control agent, to facilitate the coordination among these intersections or agents. However, they acknowledged that this may considerably increase the state space and make the training time of Q-Learning a serious issue.

Bingham proposed an isolated intersection traffic control strategy based on a generalized approximate reasoning-based intelligent control (GARIC) algorithm (*23*) developed by Berenji and Khedkar (*29*). The GARIC algorithm is essentially an actor–critic reinforcement learning (*26*), in which there are two major components called action selection network (ASN) and action evaluation network (AEN). The ASN is in the form of a fuzzy logic controller. Given certain state of the environment, it generates an action output representing how long the current green signal should be extended. The AEN is a fully connected feed-forward neural network used to approximate the value of each state. An isolated intersection with two one-way streets is considered by Bingham in the case study. She used two state variables. The first state variable APP is the number of vehicles in the movement being given green signal. The second state variable QUE is the number of vehicles in the movement being given red signal. APP and QUE are the input to both the AEN and ASN.

Most previous reinforcement learning control studies considered intersections with one-way streets or without turning movements, which are not very common in the real world. Choy et al. (*30, 31*), and Srinivasan and Choy modeled a regional traffic signal control problem using reinforcement learning (*32*). In their studies, each intersection is controlled by a pretimed controller. Reinforcement learning is used to dynamically update the cycle lengths and other parameters in response to changing traffic flow. Recently, Xie et al. developed a neuro-fuzzy reinforcement learning method for isolated intersection and arterial traffic signal control (*24*). The authors considered realistic intersection settings. Two control strategies for fixed phasing sequence and variable phasing sequence are proposed. These two strategies are applied to four isolated intersections and an arterial based on VISSIM simulation. No explicit coordination strategies are considered for the arterial signal control. The developed neuro-fuzzy reinforcement learning control is compared with optimized pretimed and actuated control. It is found that the overall performance of the reinforcement learning method is better, especially for morning and afternoon peak hours with high traffic volumes.

Reinforcement learning is an important branch of AI and has shown promising potential in solving adaptive traffic signal control problems. It can also be used for other traffic and transportation applications such as ramp metering. There are many interesting topics in this area that deserve further exploration, including how to properly define the reward function, how to identify the best state variables, etc.

SUMMARY

The existing applications of AI in traffic signal timing and optimization are primarily based on evolutionary algorithms, fuzzy logic control, artificial neural networks, and reinforcement learning algorithms. Evolutionary algorithms have mainly been used for optimizing signal timing parameters as an alternative to exact algorithms. Some of these evolutionary algorithms, such as hill climbing and GAs, have been successfully implemented in commercial software tools and demonstrated very good performance. The applications of fuzzy logic control are primarily for developing innovative control strategies that can incorporate human experience. This area has

been extensively investigated by researchers. However, not too many real-world implementations have been found. Artificial neural networks are often used in conjunction with other methods for traffic signal timing. In some cases, they are employed to estimate the performance (i.e., MOEs) of different traffic signal timing plans generated by a separate signal timing module. Artificial neural networks have also been integrated with reinforcement learning methods for traffic signal control. Reinforcement learning is a relatively new concept in the traffic signal timing arena. A major advantage of this method is that it can learn the optimal control strategy from the interactions between the control strategy (agent) and the traffic (environment), without knowing explicitly the underlying relationship between them. Although encouraging traffic signal control results have been obtained from various reinforcement learning methods, additional research is still needed before they can be applied in practice.

AI is an exciting and dynamic research area. New ideas and improvements to existing AI methods are being introduced all the time, which keep generating additional research opportunities. It is always interesting to experiment with these new ideas and see how they can address the limitations (described throughout the review) of existing AI methods or improve the performance of traditional models. Another interesting direction is to further integrate AI with traditional control and optimization methods, utilizing their respective strengths to improve traffic signal timing and optimization.

REFERENCES

1. Weise, T. *Global Optimization Algorithms—Theory and Application*. University of Science and Technology of China, Hefei, China. http://www.it-weise.de/projects/book.pdf. Accessed Dec. 31, 2011.
2. Foy, M. D., R. F. Benekohal, and D. E. Goldberg. Signal Timing Determination Using Genetic Algorithms. In *Transportation Research Record 1365*, TRB, National Research Council, Washington, D.C., 1992, pp. 108–115.
3. Park, B., C. J. Messer, T. and Urbanik II. Traffic Signal Optimization Program for Oversaturated Conditions: Genetic Algorithm Approach. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1683*, Transportation Research Board of the National Academies, Washington, D.C., 1999, pp. 133–142.
4. Park, B., C. J. Messer, and T. Urbanik II. Enhanced Genetic Algorithm for Signal-Timing Optimization of Oversaturated Intersections. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1727*, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 32–41.
5. Ceylan, H., and M. G. H. Bell. Traffic Signal Timing Optimization Based on Genetic Algorithm Approach, Including Drivers' Routing. *Transportation Research Part B: Methodological*, Vol. 38, No. 4, 2004, pp. 329–342.
6. Sun, D., Benekohal, R. F. and Waller, S. T. Multiobjective Traffic Signal Timing Optimization Using Nondominated Sorting Genetic Algorithm. In *Proceedings of the 2003 IEEE Intelligence Vehicles Symposium*, 2003, pp. 198-203.
7. Lee, J., B. Abdulhai, A. Shalaby, and E. H. Chung. Real-Time Optimization for Adaptive Traffic Signal Control Using Genetic Algorithms. *Journal of Intelligent Transportation Systems*, Vol. 9, No. 3, 2005, pp. 111–122.
8. Kennedy, J., and R. Eberhart. Particle Swarm Optimization. *Proc., IEEE International Conference on Neural Networks*, Vol. 4, 1995, pp. 1942–1948.

9.   Wei Y., Q. Shao, Y. Han, and B. Q. Fan. Intersection Signal Control Approach Based on PSO and Simulation. *Proc., 2008 Second International Conference on Genetic and Evolutionary Computing*, 2008, pp. 277–280.
10. Putha, R., L. Quadrifoglio, and E. Zechman. Comparing Ant Colony Optimization and Genetic Algorithm Approaches for Solving Traffic Signal Coordination under Oversaturated Conditions. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 27, No. 1, 2012, pp. 14–28.
11. Pappis, C. P., and E. H. Mamdani. A Fuzzy Logic Controller for a Traffic Junction. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 7, No. 10, 1977, pp. 707–717.
12. Chiu, S., and S. Chand. Adaptive Traffic Signal Control Using Fuzzy Logic. *Proc. 2nd IEEE International Conferences on Fuzzy Systems*, Thousand Oaks, California, 1993, pp. 1371–1376.
13. Niittymaki, J., and M. Pursula. Signal Control Using Fuzzy Logic. *Fuzzy Sets and Systems*, Vol. 116, No. 1, 2000, pp. 11–22.
14. Zhang, L., H. Li, and P. D. Prevedouros. Signal Control for Oversaturated Intersections Using Fuzzy Logic. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, D.C., 2005.
15. Trabia, M. B., M. S. Kaseko, and M. Ande. A Two-Stage Fuzzy Logic Controller for Traffic Signals. *Transportation Research Part C*, Vol. 7, No. 6, 1999, pp. 353–367.
16. Murat, Y. S., and E. Gedizlioglu. A Fuzzy Logic Multi-Phased Signal Control Model for Isolated Junctions. *Transportation Research Part C*, Vol. 13, No. 1, 2005, pp. 19–36.
17. Niittymaki, J. Installation and Experiences of Field Testing a Fuzzy Signal Controller. *European Journal of Operational Research*, Vol. 131, 2001, pp. 273–281.
18. Nakatsuji, T., and T. Kaku. Development of a Self-Organizing Traffic Control System Using Neural Network Models. In *Transportation Research Record 1324*, TRB, National Research Council, Washington, D.C., 1991, pp. 137–145.
19. Gilmore, J. F., and N. Abe. Neural Network Models for Traffic Control and Congestion Prediction. *Journal of Intelligent Transportation Systems*, Vol. 2, No. 3, 1995, pp. 231–252.
20. Hua, J., and A. Faghri. Development of Neural Signal Control System—Toward Intelligent Traffic Signal Control. In *Transportation Research Record 1497*, TRB, National Research Council, Washington, D.C., 1995, pp. 53–61.
21. Saito, M., and J. Z. Fan. Artificial Neural Network–Based Heuristic Optimal Traffic Signal Timing. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 15, No. 4, 2000, pp. 293–307.
22. List, G. F., and M. Cetin. Modeling Traffic Signal Control Using Petri Nets. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, No. 3, 2004, pp. 177–187.
23. Bingham, E. Reinforcement Learning in Neurofuzzy Traffic Signal Control. *European Journal of Operational Research*, Vol. 131, No. 2, 2001, pp. 232–241.
24. Xie, Y., Y. Zhang, and L. Li. Neuro-Fuzzy Reinforcement Learning for Adaptive Intersection Traffic Signal Control. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.
25. *Transportation Research Circular E-C014: Traffic Analysis Software Tools*. TRB, National Research Council, Washington, D.C., 2000. Available at http://onlinepubs.trb.org/onlinepubs/circulars/ ec014.pdf. Accessed Dec. 31, 2011.
26. Sutton, R. S., and A. G. Barto. Reinforcement Learning: An Introduction. The MIT Press, Cambridge, Mass., 1998.
27. Thorpe, T. L. Vehicle Traffic Light Control Using SARSA. Department of Computer Science, Colorado State University, Fort Collins. Available at www.cs.colostate.edu/~anderson/pubs/ thorpems.ps.gz. Accessed Dec. 31, 2011.
28. Abdulhai, B., R. Pringle, and G. J. Karakoulas. Reinforcement Learning for True Adaptive Traffic Signal Control. *Journal of Transportation Engineering*, Vol. 129, No. 3, 2003, pp. 278–285.
29. Berenji, H. R., and P. Khedkar. Learning and Tuning Fuzzy Controllers through Reinforcements. *IEEE Transactions on Neural Networks*, Vol. 3, No. 5, 1992, pp. 724–740.

30. Choy, M. C., D. Srinivasan, and R. L. Cheu. Cooperative, Hybrid Agent Architecture for Real-Time Traffic Signal Control. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, Vol. 33, No. 5, 2003, pp. 597–607.
31. Choy, M. C., R. L. Cheu, D. Srinivasan, and F. Logi. Real-Time Coordinated Signal Control Through Use of Agents with Online Reinforcement Learning. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1836*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 64–75.
32. Srinivasan, D., and M. C. Choy. Cooperative Multi-Agent System for Coordinated Traffic Signal Control. *IEE Proceedings—Intelligent Transport Systems*, Vol. 153, No. 1, 2006, pp. 41–50.

Short-Term Traffic Prediction

The next two articles are dedicated to discussing the topic of artificial intelligence (AI) and short-term traffic prediction. The first article, by J. W. C. van Lint and C. P. I. J. van Hinsbergen, provides a taxonomy of the many different approaches reported in the literature for the general problem of short-term traffic prediction; in these cases, AI techniques are discussed either as a complete solution or as part of a hybrid approach to short-term prediction. The second article, by Ghassan Abu-Lebdeh, focuses on a specific application of an AI paradigm through a case study that uses neural networks for travel time estimation and prediction on arterials.

Short-Term Traffic and Travel Time Prediction Models

Hans (J. W. C.) van Lint
Chris (C. P. I. J.) van Hinsbergen
Delft University of Technology

Road traffic is the visible result of the complex interplay between traffic demand (the amount and mix of vehicles arriving at a particular place and time) and traffic supply (e.g., capacity, prevailing speeds, and other average traffic characteristics). As a result, short-term prediction of road traffic variables is a complex nonlinear task that has been the subject of many research efforts in the past few decades. The term "short term" usually entails that the variables of interest are predicted for a period up to 1 h ahead, although the exact definition differs largely between approaches. In practical terms, short-term traffic prediction is an important if not critical component for intelligent transportation systems (ITS) and particularly in traffic control and traffic information provision.

TAXONOMY OF SHORT-TERM TRAFFIC PREDICTION METHODS

Before we discuss the different prediction approaches, we will first briefly discuss some general aspects related to short-term traffic prediction.

Some Preliminary Notes on Short-Term Traffic Prediction Approaches

Traffic prediction approaches differ largely from one another in many different aspects. These include, for example, the prediction method (further taxonomy is provided below); the scale of prediction (a fixed location, a route, or a whole network); and the type of traffic road or network involved (isolated locations, urban or freeway, controlled or not). Arguably the most important factor in assessing and comparing traffic prediction approaches is the type of input–output problem solved, that is, which traffic variables are predicted and which data are used to construct and calibrate the method do so. In very general terms the short-term traffic prediction problem boils down to solving the following regression problem:

$$y_k = G\left(x_{s|s<k}, e_{s|s<k}, \Theta_k\right) + e_k$$

(1)

where
G = the chosen model [artificial intelligence (AI) or otherwise];
y_k = (a vector of) the output variable(s) at discrete time period k;
x_k = (a vector of) the input variable(s) that, of course, could also contain historical instances of y_k;
Θ_k = (a vector of) adjustable parameter(s); and
e_k = a noise process that represents factors we cannot observe.

The general rule used in many AI textbooks, is that AI approaches (e.g., neural networks) are good candidates for the mapping G in cases where (*a*) the input–output data $\{x_t, y_t\}$ are noisy; (*b*) the relationships between these variables are multivariate and (highly) nonlinear; and (*c*) the mapping or relationship is poorly understood, that is, there are no valid (physical) theories or models to explain it. For prediction specifically, the key difficulty is in finding a balance between a sufficiently sophisticated and complex model that still is general and smooth enough to make valid forecasts given the available data [which is often referred to as the principle of parsimony or an example of Ockham's famous razor (*1, 2*)].

Given the fact that particularly the first two criteria are usually met in traffic prediction problems, applying AI techniques for the short-term prediction of traffic variables, seems—a priori—a natural choice. Suppose the (output) variable of interest is, for example, the flow or volume [vehicles per hour (vph)] on some isolated location and all (input) data available are historical time series of volumes at that same location, then clearly a time series type of approach is a sensible solution, ranging from simple auto regressive (AR) or moving average (MA) approaches to more complex integrated and possibly nonlinear (AI) approaches (*3–9*). If the desired output variable is again volume, but this time measured in a dense traffic network with other (input) traffic variables on the same or adjacent links available (perhaps including control and other relevant information), the choice of an appropriate method is much more involved. Clearly, in this case knowledge of traffic flow operations over both space and time (e.g., shockwave theory, queuing, the effect of traffic control, arrival processes) is critically important in assessing whether multivariate regression, or even more complex nonparametric (e.g., AI) approaches are sensible solutions, or that some sort of traffic simulation approach is the better choice. There is a wide range of parameterized physical models available that, in conjunction with data assimilation techniques [e.g., extended or ensemble Kalman filters (KF)], provides alternative solutions to traffic prediction, particularly over traffic corridors and entire networks (*10–12*). The main benefit of such "white box" approaches is that these models can also be used for decision support, scenario analysis, and even real-time control. Nonetheless, physical models also have their limitations in predicting traffic conditions, most importantly because of the fact that they, in turn, require predicted inputs such as demands (volumes) at the boundaries, turn fractions, or route choice patterns, and many other often poorly observable parameters. For these boundary prediction problems, again AI approaches provide excellent candidates.

A second preliminary note relates to which statistics and measures of effectiveness are used to train and assess different approaches to travel time prediction. These range from correlation coefficients, the root mean square error (RMSE), mean absolute percentage error (MAPE), and any other statistic expressing the distance between model predictions and observations. Most commonly RMSE and MAPE are used. RMSE results are insightful since they have the same unit as the predicted quantity (e.g., travel time or speed). Both measures assess results irrespective of the absolute magnitude of the prediction, which in turn makes a comparison difficult. A MAPE of 50% for the prediction of a traffic volume of 5 vph is not necessarily a bad result (e.g., in case of Poisson arrivals); a similar score for a traffic volume close to freeway capacity (say 2,000 vph) would be considered very bad.

Taxonomy of Traffic Prediction Approaches and the Role of AI Techniques

There are many overviews or comparative studies available in the literature, for example Arem et al. (*13*), Huisken and Maarseveen (*14*), Lee et al. (*15*), Lin, Zito, and Taylor (*16*), Nikovski et al.

(*17*), Ran (*18*), Smith and Demetsky (*19*), Vlahogianni, Golias, and Karlaftis (*20*), You and Kim (*21*), and Van Hinsbergen, van Lint, and Sanders (*22*), which in many cases discuss a selection of the aspects listed above. In this TRB circular we will classify traffic prediction models using the taxonomy proposed in Van Hinsbergen, van Lint, and Sanders (*22*), which differentiates the literature on the basis of the type of method used for traffic prediction (Figure 1). Due to the broad ground covered, the methods are described only qualitatively. Three (overlapping) strands of traffic prediction approaches can be identified in Figure 1.

Naïve approaches, where neither model structure (*G* in Equation 1) nor model parameters Θ_k are deduced from (real-time) data. The term naïve is rather subjective, but can be loosely interpreted as "with no additional assumptions" other than the data used and exact physical relationships (e.g., distance = speed × time). Naïve methods are widely applied in practice because of their low computational effort and easy implementation, but the accuracy is usually low. Examples include instantaneous, measured travel time or a historical average. There are also prediction approaches that combine naïve methods with nonparametric approaches such as regression or clustering (*17, 23, 24*).

Parametric (or model-based) approaches employ analytical models (e.g., travel time functions, queuing models) or traffic simulation models (microscopic or macroscopic) with parameters (link capacities other fundamental diagram parameters or car-following or lane-changing parameters) and inputs [inflows, origin–destination (O-D) flows, turn fractions, etc.] tuned to real-time data. In this case the structure of *G* in Equation 1 is fixed. Some of these parameters (Θ_k) may be freely adjusted whereas other parameters are tuned within physical bounds (e.g., capacities and free speeds must adhere to what is physically possible). Although these models implement (plausible) theoretical or physical assumptions on the time evolution of traffic (queuing, car following, shockwave theory, etc), and may be calibrated to reproduce many traffic situations, these model require data assimilation techniques to "synchronize" the internal model variables (densities or other time-varying parameters or inputs) with the real-life data. The most popular are (extended) KFs (or one of their many variations) (*10, 25–27*). Arguably, these can be classified as hybrid traffic prediction approaches that combine both parametric and nonparametric (online estimation) techniques (*28*).
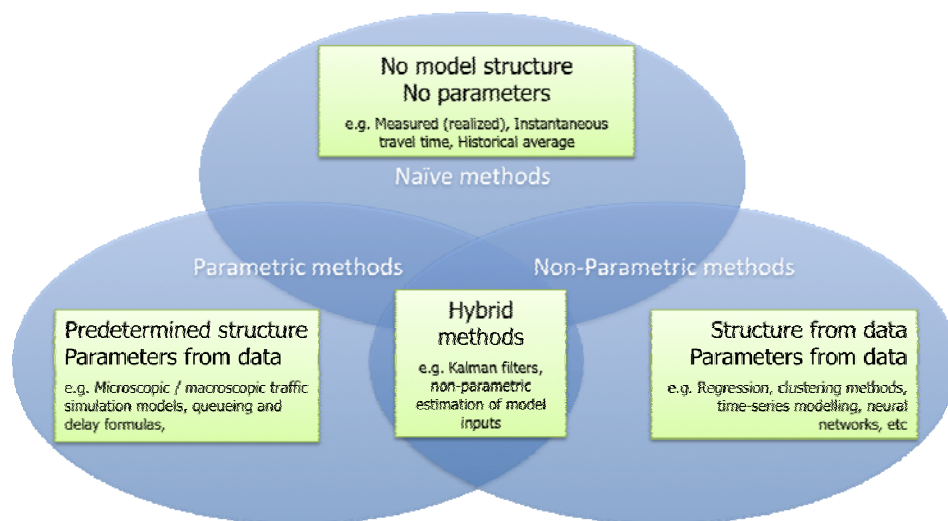
FIGURE 1 Taxonomy of traffic–travel time prediction models.

Nonparametric approaches essentially encompass all other traffic prediction methods. Typically they use generic (flexible) mathematical structures with adjustable parameters. Arguably, the largest number of traffic and travel time prediction models fall under this category. Examples include support vector regression approaches (*29*), generalized linear regression (*30, 31*), seasonal ARIMA approaches (*4*), nonlinear time-series (*7*), state–space models and KFs (*32, 33*), feed forward neural networks (*34, 35*), and recurrent neural networks (*36*), to name a few.

In the remainder of this section we will overview the literature according to this taxonomy with some examples and many references in each category. Since traffic flow theory and simulation is the subject of several other TRB committees (most importantly, the AHB45 TRB Committee on Traffic Flow Theory and Characteristics), we will not overview the broad (and fast-growing) category of parameterized and hybrid approaches to traffic prediction other then provide a few links and references and some general remarks.

NAIVE METHODS

The term naïve is rather subjective, but can be interpreted as "without any model assumption". Naïve methods are widely applied in practice because of their low computational effort and easy implementation. Strictly speaking, also the use of the last measure traffic variable (speed, flow, travel time) as a proxy for the next speed, flow or (departure) travel time can be understood as a naïve prediction method, although in this case there is no computing involved whatsoever.

Instantaneous Travel Time

The basic assumption here is that the prevailing traffic conditions (speeds, densities, queues, etc.) will remain constant indefinitely. In that case, well-understood travel time reconstruction techniques can be used as a proxy for departure travel time over a route consisting of consecutive sections i, through $TT_k = \Sigma_i L_i / u_i$. Instantaneous travel will provide accurate predictions in cases in which traffic conditions are literally stationary and homogeneous over longer time periods, e.g., in free-flowing conditions or in the rare case of a virtually fixed length queue that dissipates at a constant speed. Although this method is very fast and many practitioners understand the reasoning, its predictive performance usually deteriorates quickly when traffic conditions move from free flow to congested conditions and back.

In many cases instantaneous travel time is used merely as a baseline predictor against which other approaches are compared (*17, 24, 35, 37–42*); we know of no examples where instantaneous travel time outperforms more intelligent approaches. Nonetheless, if we consider the "state-of-the-practice" (in the Netherlands, France, Germany, England and many other countries), instantaneous travel time is still the most widely used method for most en-route travel time information panels. This implies still a lot of "missionary" work needs to be done by our community to put forward more intelligent (and accurate and reliable) approaches.

Historical Averages

Also the historical average travel time is often used as a baseline predictor against which other methods are compared. For long-term traffic prediction approaches the historical average in many cases provides the best (at least the most parsimonious) approach available. In many cases

(*17, 35, 40–44*) these averages are conditioned on either a certain time of day or day of the week period or even further (incorporating seasonality). Only in cases where the historical distribution of travel time on a given time, day, or circumstance is highly peaked around some mean value and has a very similar shape throughout the day (week, year, etc.) does this naïve method provide accurate enough predictions for short-term purposes (control, information). In reality, on many road stretches the travel time distribution is both wide and skewed (*45*), which implies that the historical average usually is a poor predictor.

Combinations of instantaneous and historical average (*46–48*) significantly improve over both these naïve approaches. There is evidence that intelligent combination of current measurements and (conditional) historical averages yields consistently better results than either current measurements or historical averages alone. In an informed historical average naïve forecast performed on par with some of the traditional nonparametric regression approaches (*49*). In a Bayesian combination approach is used to combine some slightly more involved (nonparametric) methods (*50*), but the underlying message is still the same. The combination of simple methods yields better results than any of those methods alone.

Instead of a priori binning of the data (according to day, week, and seasonality, etc.), clustering methods finally are used to compute conditional historical averages on the basis of automated grouping of different historical records of traffic data which represent similar traffic patterns. Applied algorithms are for example the small large ratio and Ward's clustering *(8, 51, 52)*. Also more involved clustering techniques (e.g., Kohonen maps) are used for essentially preprocessing input data to other (nonparametric) short-term prediction techniques (*9, 49*). These clustering techniques have been shown to outperform classical historical averages and on one occasion a linear regression.

PARAMETRIC MODELS

The term "parametric" in this context relates to the fact that only the parameters of the model are fitted using data; the structure of the model is predetermined on the basis of (traffic flow) theoretical considerations. Note that most parameters also have a much smaller feasible region, since they represent actual physical phenomena, such as critical speed and density, capacity, free speed, turning fractions, etc. Since there are other communities within TRB that deal with traffic flow theory and simulation [for an overview see Hoogendorn and Bovy (*53*)], below follows just a very brief sketch of the different parametric approaches to predict traffic.

Analytical Approaches

The easiest method for predicting travel time is to use (for many traffic engineers a well-known method) analytical formulae, such as BPR functions $TT_k = \alpha[Q_k/C_k]^{\beta}$, where Q_k depicts the expected demand, C_k the capacity, and α and β are adjustable parameters; or queuing functions of the form $TT_k = TT_{\text{free}} + N_k/C_k$, in which N_k is an estimate of the number of vehicles queuing on a route/for a bottleneck. The main problem here is that these models work only if the input variables (demand) are accurately measured, which in reality is seldom the case. Moreover, all variables and parameters are highly stochastic (demand, capacity). Typically queuing models suffer from "drift", which implies that small input errors lead to cumulative and very large prediction errors. Note that these queuing methods (store-and-forward models) are used in

practice in many traffic control applications, where the predictions are used over short road stretches and over short prediction horizons (in the order of cycle times); an example is the TUC approach (*54*).

Macroscopic Traffic Flow Models

In macroscopic simulation models, average variables of traffic streams are considered, such as the densities, mean speeds, and flows. The simplest (first order) approach uses three equations: a dynamic equation for the evolution of average density ρ (conservation of vehicles), an equilibrium relation to compute the flow q (or speed u) from density (the fundamental diagram) and the continuity equation $q = \rho u$. On a network scale, route choice patterns increase the complexity and number of degrees of freedom tremendously. In a short-term prediction context usually only aggregate route choice is considered, i.e., by means of nondestination specific turn-fractions at junctions and inflows at the boundaries. Both these need to be estimated from real-time and historical data (*28*).

   To sync the internal state variables (speed. density) with actual data from loops, floating cars or otherwise, extended KFs or more involved relatives such as unscented KF and Monte Carlo–particle filtering techniques are the most popular data assimilation techniques. For examples of networkwide short-term prediction approaches based on macroscopic traffic flow models and extended KFs we refer to Wang, Papageorgiou, and Messmer *(10, 55, 57)*, Van Hinsbergen et al. (*56*), and Tampere and Immers (*58*). Again, proper setting of the many parameters (capacity and other fundamental diagram parameters) and accurate prediction of the inputs and boundary conditions are essential. Note that these methods illustrate the versatile use of the KF, which has also been applied as an online learning technique for nonparametric traffic prediction approaches such as neural networks, and is used in many different ways in other prediction methods (*27*, *32*, *59–63*) (see also further below).

Microscopic, Mesoscopic, and Other Approaches

For short-term prediction purposes macroscopic traffic flow models are a parsimonious choice, however, also more disaggregate approaches have been (successfully) proposed. Examples include Dynasmart (*64*) and DynaMIT (*12*, *65*), which both consider destination-specific route choice and employ a so-called mesoscopic approach to model or predict traffic flow operations. Another popular prediction method involves cellular automata (CA) (*66–72*), which is a discrete and also parsimonious method to represent traffic operations, in which either route choice algorithms or O-D matrices are used to feed the predictions (*67–69*) or in which aggregate turn fractions and inflows are used (*70–72*). In a multiagent system (*73*) each driver is modeled as an agent who has an individual mental state. In an agent's mind, there is information of the trip, happiness about his current state, a set of plans of for example when to leave, and so on. These models have been combined with O-D matrix predictions to come up with networkwide traffic predictions (*74*). Finally, based on Boris Kerner's three-phase traffic theory (*75*), the ASDA and FOTO models were developed. Kerner distinguishes three phases in traffic: (*a*) free flow, (*b*) synchronized flow, a congestion phase which is usually at a fixed location, and (*c*) wide-moving jams, a congestion phase that moves upstream the traffic with a constant speed. ASDA and FOTO track and predict the characteristics of these phases. Reported results show that

calculation times are lower than with microscopic simulation and that accuracy is satisfactory, although the precise algorithms employed are not disclosed in detail due to patents.

NONPARAMETRIC MODELS

The term nonparametric in our taxonomy does not imply "without parameters" but is indicative of the fact that the number and nature of the parameters are flexible and not fixed in advance (*76*). In other words, both model structure (e.g., the degree of a polynomial, the number of layers in a neural network, the distance function in unsupervised mapping techniques) and the values of the model parameters are determined from data.

Linear Time Series and Regression Approaches

These two popular families of traffic prediction approaches arise in case the function *G* in Equation 1 is linear in its parameters Θ (not necessarily in its arguments *x,* e.g., also polynomials of the general form $y = \Sigma_m[a_m x^m]$ fall in this category). The key to successful applying these approaches to traffic prediction is that the output time series can indeed be approximated and reproduced by weighted linear combinations of (possible transformed or preprocessed) input time series.

Time series prediction involves modeling a variable y_k as a parameterized (weighted) linear function of past observations y_{k-n} of that variable (AR) and past error terms e_k (MA processes). They are essentially a special case of regression models, in which the input $x_{s|s<k}$ is composed of past observations of *y*. ARMA models assume that the process generating the data is stationary (i.e., there is no structural trend in the time series). Since traffic (and many other real-world) processes are typically not stationary, usually a term for structural trends in the data is integrated. The resulting model is generally referred to as an ARIMA (*p*, *d*, *q*) model where *p*, *d*, and *q* are nonnegative integers that refer to the order of the autoregressive, integrated (trend), and moving average parts of the model respectively. In the seminal book of Box and Jenkins (*77*) a still widely used structured methodology is put forward to determine these terms and their weights from data. An early overview of ARIMA models applied to traffic forecasting can be found in (*78*), and since then many examples are available in the literature (*4, 43, 48, 78–82*). Additionally, many variations or additions on ARIMA have been proposed in literature, such as seasonal ARIMA (SARIMA) (*4, 83, 84*) in which periodic terms are added (which typically pertain to time of day, day of week, or other trends in traffic data); subset ARIMA models, which partition the time series in subsets with corresponding terms and components (*81*); variations on exponential smoothing (*49, 85*); models in which more sophisticated unsupervised techniques are used in determining the weights and terms such as Kohonen ARIMA (*9*); models in which also exogenous inputs (from other time series) are incorporated (ARIMAX) (*86*). These could also be understood as special cases of regression models; and many other methods such as vector and spatiotemporal [VAR(I)MA, STAR(I)MA] models (*43*).

It is important to note that not all traffic prediction problems schematized by Equation 1 can be cast into a time series problem. Most importantly, ARIMA approaches do not apply to departure time prediction (*59*). The reason is that with time periods *k* of fixed size Δt (of typically 1, 5, or 10 min in short-term prediction problems), there is no guarantee that a previous departure travel time TT_{k-n} is available for input, with *n* typically depicting 1 or a few discrete

time periods. In fact TT_{k-n} is actually measured at time period $k^* = k - n + TT_{k-n}$, which in most nontrivial cases will be later than k. This problem becomes rapidly worse when one considers a reasonable sized (e.g., more than a few kilometers) and congested route with travel times typically (much) larger than Δt. The implication is that time series approaches to travel time prediction only work offline (when the data are available) but not in a real (online) traffic information or control applications; and that, counterintuitively, when predicting departure travel times on longer routes, measured (by definition arrival) travel times are of little use as inputs (*59*). Arguably one could predict arrival travel times (which are available at every time step) with an ARIMA process, but these are of little more use in providing traffic information or traffic control than measured arrival times. Of course, measured travel times are quintessential as desired outputs during calibration or training of virtually all traffic prediction models, and also for many other purposes.

For travel time prediction, and also the prediction of traffic variables on locations where both in space and time additional (traffic) data are available (and needed), linear regression approaches are a more suitable alternative. In regression models prediction function is assumed to be a linear combination of its covariates, where parameters indicate how much one covariate contributes to the outcome (*17, 49, 87–93*).

Variations on Regression and Time Series Modeling

The **KF** is a recursive minimum variance estimator that guarantees an optimal solution for all linear problems that are cast in state space form (*94*); in the same general form as Equation 1, this reads as

$$\mathbf{z}_k = G\left(\mathbf{z}_{k-1}, \mathbf{x}_{k-1}, \mathbf{e}_{k-1}^{(x)}, \Theta_k^{(x)}\right)$$
$$\mathbf{y}_k = H\left(\mathbf{z}_k, \mathbf{e}_k^{(y)}, \Theta_k^{(y)}\right)$$

(2)

The top (process) equation in Equation 2 computes the evolution of the state z; the bottom (observation) equation relates the observable outputs to this state. Since with a bit of creativity virtually all regression and time series problems can be cast in state–space form, the KF is a very powerful tool in many estimation and prediction problems. One of its appealing properties is that it requires no memory, i.e., it predicts the future state z_k of a system using the last available state z_{k-1} and whatever new information (x_{k-1}) is available. This makes it very suitable to apply in online estimation or calibration settings. Even neural networks and other nonlinear (AI) traffic predictors can be trained using the (extended) KF, in which case the parameters represent the unobserved state, which is assumed to follow a random walk (the simplest of MA models) and the neural network itself is assumed to make a (nonlinear) observation of the state (i.e., the neural network is represented by observation model H). Examples of linear KF filters applied to traffic prediction include (*61, 63, 82, 95–102*) Examples of nonlinear (i.e., extended) Kalman filtering applied to traffic prediction include (*10, 27, 32, 55–63*). In the latter nonlinear case, there is no guarantee that the estimator is optimal.

Another popular approach is **locally weighted regression**. The prediction residual of each data point is then weighted proportionally to its proximity to the current measurement. Very good results are reported (*17, 93*) in prediction accuracy as well as computation time.

Further examples of prediction models in this broad category include amongst others the **ATHENA** model. In this model traffic is modeled as a linear combination of historical and

current states. For each type of traffic a nonlinear transformation is applied (*103*). This model outperformed several other models (*80*). As a final example we refer to the so-called self-exciting threshold AR (**SETAR**) model, which uses a linear combination of the current measurement and one past measurement to predict the future state (*104–106*). The current measurement is weighted more heavily. Although this model is fast, the accuracy is low.

Artificial Neural Networks

Artificial neural networks (ANNs) are the most widely applied models to the traffic prediction problem. As noted above, for traffic prediction purposes, ANNs can be understood as nonlinear regression models, although ANNs also are used in this context for clustering, classification, and feature extraction. Note that a linear regression model of the form $\mathbf{y} = \theta_0 + \sum_{i=1}^{M} \theta_i \mathbf{x}_i$ represents the simplest ANN conceivable, i.e., one with a single neuron, *M* inputs, and a linear transfer function (the identity function *G*: $z \rightarrow z$). The most commonly used ANN is the multilayer perceptron (MLP). Mathematically, a single output MLP model with two layers, *M* inputs and *N* hidden neurons reads

$$y = h\left(\varphi_0 + \sum_{j=1}^{N} \varphi_j g\left(\theta_0 + \sum_{i=1}^{M} \theta_i \mathbf{x}_i \right) \right) \qquad (3)$$

in which *h* and *g* are nonlinear functions (at least *g* is) of their arguments. With such a model a wide range of regression and time series problems can be accommodated. The simplest method of fitting the parameters $\Theta = [\boldsymbol{\varphi}, \boldsymbol{\theta}]$ of this model to data in the least squares sense (e.g., minimizing the sum of squared errors) is error back propagation (BP), which uses the chain rule to relate output errors $\varepsilon_k = y_k - y_k^{obs}$ to changes in the internal weights. Many studies still use BP neural networks (BPNNs) to predict traffic data (*6, 24, 47, 85, 107–116*) and do so with varying results, although almost in all cases, more sophisticated models and training, improve upon the results.

Improved ANN Training

The main problem with BP is that is a gradient-based method that uses local information (the gradient of the error function with respect to the weights) and as such leads to a local minimum that best fits the available training data. That does not guarantee at all (often on the contrary) that the trained model can make plausible predictions. Besides the obvious necessity of assessing the predictive performance of these models to different data than the ones used for training (i.e., use of validation and test data sets), many better and faster converging training methods have been developed since back propagation was popularized in the mid 1980s by Rumelhart (*117*) amongst others. Conjugate gradient algorithms (CGAs) use higher order information (on the local curvature of the error function) to determine the magnitude and direction of weight adaptation. The most widely known CGA is the Fletcher–Reeves update (*118, 119*). Another popular alternative is the Levenbergh–Marquardt method (*120, 121*), which is an approximation of Newton's method that also uses second order information to determine stepsize and direction of weight updates. Evolutionary learning (applying genetic algorithms) provides an alternative approach to (global) training and designing neural networks. These evolutionary neural networks

show very positive results (*89*, *93*, *122–125*). Finally, important steps forward have been made in preventing ANNs from overfitting the data and assuring their predictive validity. These methods include cross-validation, bootstrapping, and random subsampling (*120*) and Bayesian techniques (*126–128*), which automatically balance the model fit to data with model complexity (i.e., number and absolute size of parameters).

Different ANN Topologies

For traffic prediction many different functional forms of ANNs (i.e., different mathematical specification of Equation 3) have been applied. The modular neural network (MNN) is based on a divide-and-conquer strategy (*110*). The input is processed in several subnetworks, each specializing in a certain task. MNNs are faster to train and can improve results (*107*, *112*, *128*, *129*). Radial basis frequency networks (RBFNN) use a hidden layer of so-called basis functions to cluster the input space, with each cluster represented by a hidden neuron. Results show a slightly positive preference of RBFNN over BPNN (*85*, *130–132*). Inside the hidden layer of a so-called neuro-fuzzy network, fuzzy rules are defined automatically (*133–135*) on the basis of data. Results are comparable to or better than those of the BPNN. In a counter propagation network, at each iteration the inputs are assigned to one node using a distance measure (*136*). Training time is dramatically decreased and performance is slightly improved. A resource allocating network is much like the RBFNN (*96*), except that hidden units are created automatically.

Differences in Preprocessing Input Data

So-called wavelet neural network use wavelet functions instead of the standard sigmoid function used in BPNN (*132*) which reportedly lead to improvements in prediction accuracy as well as computational effort (*137*, *138*). The Spectral Basis Network (SNN) employs a Fourier expansion of the input vector to obtain linearly separable input features (*35*, *41*). This new input vector is fed to a standard BPNN model. Improvements in prediction accuracy are found, especially on longer prediction horizons. The generalized network (GNN) also uses a Fourier expansion of the input vector. The hidden neurons however are replaced with intelligent neurons which have an increased storage capability. In convergence and in accuracy, the GNN is found to be better than the BPNN in at least one study (*139*). Another alternative is to use a Kohonen self organizing feature map (SOFM) network to cluster input data before feeding it to a standard BPNN. In one study this SOFM BPNN is found to outperform a standard BPNN, but is in its turn outperformed by a fuzzy c-means clustering network (*40*). In a principal component analysis neural network, the input vector data are compressed rather than expanded, reducing the number of inputs and hence neural network complexity, therefore improving the BPNN performance (*107*, *110*).

Dynamic Neural Networks

The Jordan–Elman network or simple recurrent network (SRNN) contains memory units that are used to store the hidden layer output signals at the previous time step, providing a mechanism to recognize recurring patterns (*107*, *140*). Essentially these SRNNs can be understood as nonlinear multivariate MA-type models, in that they estimate the contribution of some unknown noise

process to the time evolution of the variables of interest. The Partially Recurrent Network (PRNN) is a simplified version of the Jordan–Elman network (*107*, *140*, *141)* and shows equal performance to other neural networks. The state–space neural network, in turn, is another variation on the Elman network (*59*, *120*, *121*, *142)*. The SSNN outperforms naïve travel time prediction methods drastically. In van Lint (*59*) it is shown that an SSNN trained with Levenbergh–Marquardt has better predictive performance than the same models trained with incremental learning algorithms. In contrast to the BPNN, in the finite impulse response network (FIRNN) the static weights are replaced by linear filters which have tapped delay lines in it, so to capture the internal dynamics of the traffic processes (*143*). The FIRNN outperforms BPNN, but is outperformed by in the same study by time delay recurrent network (TDRNN) where the previous output values are fed back into the input values (*143*, *144*). These TDRNNs can be understood as nonlinear multivariate AR models.

Other Nonlinear Time Series and Regression Approaches

There are so many methods in and around the domain of artificial intelligence applied to the traffic prediction problem, that it is impossible to treat all of them. Below follow a few additional examples. With the *k*-nearest neighbor method, a historical database is searched every time for the *k* events which are nearest to the current traffic situation. The outcomes of the nearest events are averaged or weighed to their distance to the current situation (*48*, *90*, *145–154*). All studies show that it is a fast method that can outperform naïve prediction methods, but none finds it more accurate than more advanced methods.

Fuzzy Logic

Fuzzy logic has been applied to traffic prediction on several occasions, mostly in combination with other techniques. The basic principle is that a rule base is created [a set of if–then (premise-consequence) rules related to fuzzy traffic variables], manually or automatically. An observation on the prevailing traffic situation then corresponds to one or more rules. Based on the premise and the degree of correspondence, a prediction is made. In Coufal and Turunen (*133*) and Li, Lin, and Liu (*155*), promising results are reported without a comparison to other methods. In Huisken (*131*), fuzzy models are compared to BPNN and RBFNN where the latter produce better predictions on that particular case.

Bayesian Belief Network

A Bayesian belief network (BBN), also known as causal model, is a directed graphical model that represents conditional dependencies between a set of random variables. The parameters that need to be estimated in BBN are the conditional probabilities that relate one belief state to the next. This method is applied in real life by one of the largest traffic information companies Inrix (*156*), a Microsoft spin-off company. In their method data from adjacent links are considered informative for the current and future state of other links (*3*). Comparisons with other methods are not made, for obvious (commercial) reasons.

Support Vector Regression

Finally, support vector regression (SVR) is a (popular) machine learning method where the goal is to find a function that has at most a certain threshold deviation from the actually obtained targets for all the training data and at the same time is as flat as possible (*157*). The SVR method can outperform naïve methods (*42*).

DISCUSSION, CONCLUSIONS, AND OUTLOOK

The first conclusion we can draw is that over the past two decades researchers in our field have used an overwhelmingly large number of different approaches to solve the short-term traffic prediction problem. Depending on application and design criteria, all of the discussed approaches have advantages and disadvantages. Naïve methods are fast, scalable, and easy to implement, but usually lack prediction accuracy unless these methods are refined with more involved methods (e.g., combinations of instantaneous and historical travel times). Nonparametric methods, and particularly the nonlinear ones such as neural networks, are in general more accurate and robust (to data failure) and are better suited to learn the nonlinear traffic dynamics from data. On the downside, nonparametric models (both linear and nonlinear) provide essentially location-specific black-box solutions; that is, they are not easily transferred from one application to another. Moreover, their predictive validity is constraint to situations that were visible in the data with which they were trained. Also training and calibration itself poses challenges in terms of finding the proper balance between model complexity and predictive validity. Finally, the vast majority of studies using AI methods focuses on predicting traffic variables on a single location or fixed routes on freeways. Typically, such models are well suited for automated (local) traffic control or traffic information purposes.

Only a small number of methods have been applied to networkwide traffic prediction. This has been done with historical averages, linear regression, and Bayesian networks, and parametric methods such as simulation models. The latter (parametric techniques) do incorporate the underlying traffic dynamics in their equations and structure and hence are better suited to deal with unseen traffic situations (incidents, excessive demands, changes in the network), and can be suited to incorporate, for example, traffic control. However, the number of adjustable inputs, parameters, and variables in such simulation models is usually much larger than the available (real-time) data, which makes real-time tuning of such model-based approaches a complex (and often underdetermined) problem.

The bottom line obviously is that there exists not a single best method in any situation, let alone under all possible situations. Given the many models and variations tried and tested, it is now time for synthesis. Further research on traffic prediction could focus on either developing truly new methods which outperform the present state of the art presented above, or focus on methods which help to select or combine the appropriate model for a given situation. With the large amount of methods already at hand, the second option will be very worthwhile and can be expected to produce at least interesting results in the foreseeable future. A few researchers already have made steps in the direction using Bayesian methods to combine different models. I suggest more effort is needed to bring these also under the attention of our peers.

The approach which in my opinion has a lot of potential is that elusive hybrid category in the center of Figure 1. The recent years have witnessed an increased number of professional

organizations interested in the short-term prediction problem, such as Inrix (using a completely data driving Bayesian belief approach); PTV/Systema (using a CTM-based dynamic simulation), RENAISSANCE (second-order macroscopic model and the EKF) and also Fileradar.nl (TU Delft spin off using CTM and a localized KF as their prediction Kernel). Given this, I expect large steps forward in the coming years in terms of networkwide solutions for traffic estimation and control, for both decision support as well travel time information and traffic control purposes.

REFERENCES

1. Bishop, C. M. *Neural Networks for Pattern Recognition*. Oxford University Press, United Kingdom, 1995.
2. MacKay, D. J. C. Probable Networks and Plausible Predictions: A Review of Practical Bayesian Methods for Supervised Neural Networks. *Network: Computation in Neural Systems*, Vol. 6, No. 3, 1995, pp. 469–505.
3. Zhang, C., S. Sun, and G. Yu. A Bayesian Network Approach to Time Series Forecasting of Short-Term Traffic Flows. *Proc., 7th International IEEE Conference on Intelligent Transportation Systems*, 2004.
4. Williams, B. M., and L. A. Hoel. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: A Theoretical Basis and Empirical Results. *Journal of Transportation Engineering*, Vol. 129, No. 6, 2003, pp. 664–672.
5. Williams, B. M. Multivariate Vehicular Traffic Flow Prediction: Evaluation of ARIMAX Modeling. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1776*, TRB, National Research Council, Washington D.C., 2001, pp. 194–200.
6. Huisken, G., and A. Coffa. Short-Term Congestion Prediction: Comparing Time Series with Neural Networks. Proc., Tenth International Conference on Road Transport Information and Control, No. 472, 2000.
7. Al-Deek, H. M., M. P. D. Angelo, and M. C. Wang. Travel Time Prediction with Nonlinear Time Series. Proc., Fifth International Conference on Applications of Advanced Technologies in Transportation, Reston, Va., 1998.
8. Wild, D. Short-Term Forecasting Based on a Transformation and Classification of Traffic Volume Time Series. *International Journal of Forecasting*, Vol. 13, 1997, pp. 63–72.
9. Voort, M. v. d., M. Dougherty, and S. Watson. Combining Kohonen Maps with ARIMA Time Series Models to Forecast Traffic Flow. *Transportation Research*, Vol. 4, No. 5, 1996, pp. 307–318.
10. Wang Y., M. Papageorgiou, and A. Messmer. A Real Time Freeway Network Traffic Surveillance Tool. *IEEE Transactions on Control Systems Technology*, Vol. 14, No. 1, 2006.
11. Van Lint, J. W. C., S. P. Hoogendoorn, and A. Hegyi. Dual EKF State and Parameter Estimation in Multi-Class First-Order Traffic Flow Models. *17th IFAC World Congress*, Vol. 17, 2008.
12. Ben-Akiva, M., M. Bierlaire, D. Burton, H. N. Koutsopoulos, and R. Mishalani. Network State Estimation and Prediction for Real-Time Traffic Management. *Networks and Spatial Economics*, Vol. 1, 2001, pp. 293–318.
13. Arem, B. v., H. R. Kirby, M. J. M. v. d. Vlist, and J. C. Whittaker. Recent Advances and Applications in the Field of Short-Term Traffic Forecasting. *International Journal of Forecasting*, Vol. 13, 1997, pp. 1–12.
14. Huisken, G., and M. v. Maarseveen. Congestion Prediction on Motorways: A Comparative Analysis. *Proc., 7th World Congress on Intelligent Transport Systems*, 2000.
15. Lee, S., D. Kim, J. Kim, and B. Cho. Comparison of Models for Predicting Short-Term Travel Speeds. *Proc., 5th World Congress on Intelligent Transport Systems*, 1998.

16. Lin, H.-E., R. Zito, and M. A. P. Taylor. A Review of Travel-Time Prediction in Transport and Logistics. *Proc., Eastern Asia Society for Transportation Studies,* 2005.
17. Nikovski, D., N. Nishiuma, Y. Goto, and H. Kumazawa. Univariate Short-Term Prediction of Road Travel Times. *Proc., 8th International IEEE Conference on Intelligent Transportation Systems,* 2005.
18. Ran, B. Using Traffic Prediction Models for Providing Predictive Traveller Information. *International Journal of Technology Management*, Vol. 20, No. 3/4, 2000, pp. 326–339.
19. Smith, B. L., and M. J. Demetsky. Traffic Flow Forecasting: Comparison of Modeling Approaches. *Journal of Transportation Engineering*, Vol. 123, No. 4, 1997, pp. 261–266.
20. Vlahogianni, E. I., J. C. Golias, and M. G. Karlaftis. Short-Term Traffic Forecasting: Overview of Objectives and Methods. *Transport Reviews*, Vol. 24, No. 5, 2004, pp. 533–557.
21. You, J., and T. J. Kim. Towards Developing a Travel Time Forecasting Model for Location-Based Services. *Proc., 43rd European Congress of the Regional Science Association*, 2003.
22. Van Hinsbergen, C. P. I., J. W. C. Van Lint, and F. M. Sanders. Short-Term Traffic Prediction Models. *Proc., 14th World Congress on Intelligent Transport Systems: ITS for a Better Life*, Beijing: Research Institute of Highway, Chinese Ministry of Communications, 2007.
23. Rice, J., and E. Van Zwet. A Simple and Effective Method for Predicting Travel Times on Freeways. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, No. 3, 2004, pp. 200–207.
24. Huisken, G., and E. C. van Berkum. Comparative Analysis of Short-Range Travel Time Prediction Methods. Presented at 82nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2003.
25. Zuurbier, F. S. and Van Lint, J. W. C. Traffic Network State Estimation Using Extended Kalman Filtering and DSMART. *Proc., 16th IFAC World Congress*, Delft, Netherlands, 2006.
26. Van Lint, J. W. C., and S. P. Hoogendoorn. The Technical and Economic Benefits of Data Fusion for Real-Time Monitoring of Freeway Traffic. *World Congress of Transportation Research*, Berkeley, Calif., 2007.
27. Antoniou, C., M. E. Ben-Akiva, and H. N. Koutsopoulos. Online Calibration of Traffic Prediction Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1934*, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 235–245.
28. Calvert, S. C., J. W. C. Van Lint, and S. P. Hoogendoorn. A Hybrid Travel Time Prediction Framework for Planned Motorway Roadworks. *Proc., 13th International IEEE Conference on Intelligent Transportation Systems*, 2010.
29. Wu, C.-H., C.-C. Wei, D.-C. Su, M.-H. Chan, and J.-M. Ho. Travel Time Prediction with Support Vector Regression. *Proc., 2003 IEEE Conference on Intelligent Transportation Systems,* Shanghai, China, 2003.
30. Sun, H., H. X. Liu, H. Xiao, R. R. He, and B. Ran. Use of Local Linear Regression Model for Short-Term Traffic Forecasting. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1836*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 143–150.
31. Zhang, X., and J. A. Rice. Short-Term Travel Time Prediction. *Transportation Research Part C: Emerging Technologies*, Vol. 11, No. 3–4, 2003, pp. 187–210.
32. Chien, S. I. J., and C. M. Kuchipudi. Dynamic Travel Time Prediction with Real-Time and Historic Data. *Journal of Transportation Engineering*, Vol. 129, No. 6, 2003, pp. 608–616.
33. Stathopoulos, A., and M. G. Karlaftis. A Multivariate State Space Approach for Urban Traffic Flow Modeling and Prediction. *Transportation Research Part C: Emerging Technologies*, Vol. 11, No. 2, 2003, pp. 121–135.
34. Park, D., L. Rilett, and G. Han. Spectral Basis Neural Networks for Real-Time Travel Time Forecasting. *Journal of Transportation Engineering*, Vol. 125, No. 6, 1999, pp. 515–523.
35. Rilett, L. R., and D. Park. Direct Forecasting of Freeway Corridor Travel Times Using Spectral Basis Neural Networks. In *Transportation Research Record: Journal of the Transportation Research*

Board, No. 1752, Transportation Research Board of the National Academies, Washington, D.C., 2001, pp. 140–147.

36. van Lint, J. W. C. Online Learning Solutions for Freeway Travel Time Prediction. *IEEE Transactions in Intelligent Transportation Systems*, Vol. 9, No. 1, 2008, pp. 38–47.

37. van Lint, J. W. C., S. P. Hoogendoorn, and H. J. van Zuylen. State space neural networks for freeway travel time prediction. Presented at 81st Annual Meeting the of Transportation Research Board, Washington, D.C., 2002.

38. van Lint, J. W. C., S. P. Hoogendoorn, and H. J. van Zuylen. Freeway Travel Time Prediction with State–Space Neural Networks: Modeling State–Space Dynamics with Recurrent Neural Networks. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1811*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 30–39.

39. van Lint, H., and M. Schreuder. Travel Time Prediction for Variable Message Sign Panels: Results and Lessons Learned from Large-Scale Evaluation Study in the Netherlands. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006.

40. Park, D., and L. R. Rilett. Forecasting multiple-period freeway link travel times using modular neural networks. In *Transportation Research Record*, Vol. 1617, 1998, pp. 163–170.

41. Park, D., L. R. Rilett, and G. Han. Spectral Basis Neural Network for Real-Time Travel Time Forecasting. *Journal of Transportation Engineering*, Vol. 125, No. 6, 1999, pp. 515–523.

42. Wu, C.-H., J.-M. Ho, and D. T. Lee. Travel-Time Prediction with Support Vector Regression. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, No. 4, 2004, pp. 276–281.

43. Kamarianakis, Y., and P. Prastacos. Forecasting Traffic Flow Conditions in an Urban Network: Comparison of Multivariate and Univariate Approaches. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1857*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 74–84.

44. Eglese, R., W. Maden, and A. Slater. A Road Timetable to Aid Vehicle Routing and Scheduling. *Computers & Operations Research*, Vol. 33, 2006, pp. 3508–3519.

45. van Lint, J. W. C., H. J. van Zuylen, and H. Tu. Travel Time Unreliability on Freeways: Why Measures Based on Variance Tell Only Half the Story. *Transportation Research Part A: Policy and Practice*, Vol. 42, No. 1, 2008, pp. 258–277.

46. Hobeika, A. G., and C. K. Kim. Traffic-Flow-Prediction Systems Based on Upstream Traffic. *Proc., 1994 Vehicle Navigation and Information Systems Conference*, IEEE, 1994.

47. Park, D., and L. R. Rilett. Forecasting Freeway Link Travel Times with a Multilayer Feedforward Neural Network. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 14, 1999, pp. 357–367.

48. Smith, B. L., B. M. Williams, and R. K. Oswald. Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting. *Transportation Research Part C*, Vol. 10, 2002, pp. 303–321.

49. Chrobok, R., O. Kaumann, J. Wahle, and M. Schreckenberg. Different Methods of Traffic Forecast Based on Real Data. *European Jorunal of Operational Research,* Vol. 155, 2004, pp. 558–568.

50. van Hinsbergen, C. P. I., and J. W. C. van Lint. Bayesian Combination of Travel Time Prediction Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2064*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 73–80.

51. Chung, E. Classification of Traffic Pattern. *Proc., 10th ITS World Congress,* 2003.

52. Weijermars, W., and E. von Berkum. Analyzing Highway Flow Patterns Using Cluster Analysis. *Proc., IEEE Conference on Intelligent Transportation Systems*, 2005.

53. Hoogendoorn, S. P., and P. H. L. Bovy. State-of-the-Art of Vehicular Traffic Flow Modeling. *Proc., Institution of Mechanical Engineers*, Vol. 215, No. 1, 2001, pp. 283–303.

54. Aboudolas, K., M. Papageorgiou, A. Kouvelas, and E. Kosmatopoulos. A Rolling-Horizon Quadratic-Programming Approach to the Signal Control Problem in Large-Scale Congested Urban Road Networks. *Transportation Research Part C: Emerging Technologies*, Vol. 18, No. 5, 2010, pp. 680–694.

55. Wang, Y., M. Papageorgiou, and A. Messmer. RENAISSANCE: A Unified Macroscopic Model-Based Approach to Real-Time Freeway Network Traffic Surveillance. *Transportation Research Part C: Emerging Technologies*, Vol. 14, No. 3, 2006, pp. 190–212.

56. van Hinsbergen, C. P. I., T. Schreiter, F. S. Zuurbier, J. W. C. van Lint, and H. J. van Zuylen. Fast Traffic State Estimation with the Localized Extended Kalman Filter, 2010.

57. Wang, Y., M. Papageorgiou, and A. Messmer. Real-Time Freeway Traffic State Estimation Based on Extended Kalman Filter: Adaptive Capabilities and Real Data Testing. *Research Part A: Policy and Practice*, Vol. 42, No. 10, 2008, pp. 1340–1358.

58. Tampere, C., and B. Immers. An Extended Kalman Filter Application for Traffic State Estimation Using CTM with Implicit Mode Switching and Dynamic Parameters. *Proc., 2007 IEEE Intelligent Transportation Systems Conference,* Seattle, Wash., 2007, pp. 209–216.

59. van Lint, J. W. C. Online Learning Solutions for Freeway Travel Time Prediction. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 9, No. 1, 2008, pp. 38–47.

60. Liu, H., J. W. C. van Lint, H. J. van Zuylen, and K. Zhang. Two Distinct Ways of Using Kalman Filters to Predict Urban Arterial Travel Time. *Proc., IEEE ITSC 2006*, Toronto, 2006.

61. Yang, J.-S. Travel Time Prediction Using the GPS Test Vehicle and Kalman Filtering Techniques. *Proc., 2005 American Control Conference*, 2005.

62. Park, D., and L. R. Rilett. Forecasting Multiple-Period Freeway Link Travel Times Using Modular Neural Networks. In *Transportation Research Record 1617*, TRB, National Research Council, Washington, D.C., 1998, pp. 163–170.

63. Okutani, I., and Y. J. Stephanedes. Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory. *Transportation Research Part B*, Vol. 18, No. 1, 1984, pp. 1–11.

64. Mahmassani, H. S. DynaSMART-X Home. 2004. Available at http://www.dynasmart.umd.edu/ dynasmartx/index.htm.

65. Ben-Akiva, M. E., N. Bierlaire, D. Burton, H. N. Koutsopoulos, and R. G. Mishalani. Network State Estimation and Prediction for Real-Time Transportation Management Applications. Presented at 81st Annual Meeting of the Transportation Research Board, Washington, D.C., 2002.

66. Nagel, K., and M. Schreckenberg. A Cellular Automaton Model for Freeway Traffic. *Journal de Physique I*, Vol. 2**,** 1992, pp. 2221–2229.

67. Chopard, B., A. Dupuis, and P. O. Luthi. A Cellular Automata Model for Urban Traffic Andits Application to the City of Geneva. *Proceedings of Traffic and Granular Flow*, 1997.

68. Dailey, D., and N. Taiyab. A Cellular Automata Model for Use with Real Freeway Data. TRAC/WSDOT Final Technical Report WA-RD 537.1. Washington State Transportation Center, 2002.

69. Kanoh, H., T. Furukawa, S. Tsukahara, K. Hara, H. Nishi, and H. Kurokawa. Short-Term Traffic Prediction Using a Fuzzy C-Means and Cellular Automata in a Wide-Area Road Network. *Proc., 8th International IEEE Conference on Intelligent Transportation Systems*, 2005.

70. Chrobok, R., O. Kaumann, J. Wahle, and M. Schreckenberg. Three Categories of Traffic Data: Historical, Current, and Predictive. *Proc., 9th IFAC Symposium "Control in Transportation Systems"*, 2000.

71. Chrobok, R., A. Pottmeier, S. F. Hafstein, and M. Schreckenberg. Traffic Forecast in Large Scale Freeway Networks. *International Journal of Bifurcation and Chaos*, Vol. 14, No. 6, 2004, pp. 1995–2004.

72. Chrobok, R., J. Wahle, and M. Schreckenberg. Traffic Forecast Using Simulations of Large-Scale Networks. *Proc., 4th International IEEE Conference on Intelligent Transportation Systems*, 2001.

73. Wahle, J., and M. Schreckenberg. A Multi-Agent System of On-Line Simulations Based on Real-World Traffic Data. *Proc., Hawaii International Conference on System Science*, IEEE, 2001.

74. Miska, M. P. Real-Time Traffic Management by Microscopic Online Simulation. Delft University of Technology, 2007.

75. Kerner, B. S. *The Physics of Traffic*. Springer-Verlag, 2004.

76. Wikipedia. Available at http://en.wikipedia.org/wiki/Nonparametric. Accessed January 3, 2007.

77. Box, G. E. P., and G. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, 1976.

78. Nihan, N. L. Use of the Box and Jenkins Time Series Technique in Traffic Forecasting. *Transportation*, Vol. 9, 1980, pp. 125–143.

79. Hamed, M. M., H. R. Al-Masaeid, and Z. M. B. Said. Short-Term Prediction of Traffic Volume in Urban Arterials. *Journal of Transportation Engineering*, Vol. 121, No. 3, 1995, pp. 249–254.
80. Kirby, H. R., S. M. Watson, and M. S. Dougherty. Should We Use Neural Networks or Statistical Models for Short-Term Motorway Traffic Forecasting? *International Journal of Forecasting*, Vol. 13, 1997, pp. 43–50.
81. Lee, S., and D. B. Fambro. Application of Subset Autoregressive Integrated Moving Average Model for Short-Term Freeway Traffic Volume Forecasting. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1678*, TRB, National Research Council, Washington, D.C., 1999, pp. 179-188.
82. Lee, Y.-I., and C. Y. Choi. Development of a Link Travel Time Prediction Algorithm for Urban Expressway. *Proc., 5th World Congress on Intelligent Transport Systems,* 1998.
83. Guo, J. Adaptive Estimation and Prediction of Univariate Vehicular Traffic Condition Series. North Carolina State University, 2005.
84. Williams, B. M., P. K. Durvasula, and D. E. Brown. Urban Freeway Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models. In *Transportation Research Record 1644*, TRB, National Research Council, Washington, D.C., 1998, pp. 132–141.
85. Park, B., C. J. Messer, and T. Urbanik II. Short-Term Freeway Traffic Volume Forecasting Using Radial Basis Function Neural Networks. In *Transportation Research Record 1651*, TRB, National Research Council, Washington, D.C., 1998, pp. 39–47.
86. Williams, B. M. Multivariate Vehicular Traffic Flow Prediction. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1776*, TRB, National Research Council, Washington, D.C., 1999, pp. 194–200.
87. Kwon, J., B. Coifman, and P. Bickel. Day-to-Day Travel Time Trends and Travel Time Prediction from Loop Detector Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1717*, TRB, National Research Council, Washington, D.C., 2000, pp. 120–129.
88. Lan, C.-J., and S.-P. Miaou. Real-time prediction of traffic flows using dynamic generalized linear models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1678*, TRB, National Research Council, Washington, D.C., 1999, pp. 168–178.
89. Lingras, P., S. Sharma, and M. Zhong. Prediction of Recreational Travel Using Genetically Designed Regression and Time-Delay Neural Network Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1805*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 16–24.
90. Rice, J., and E. van Zwet. A Simple and Effective Method for Predicting Travel Times on Freeways. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, No. 3, 2004, pp. 200–207.
91. Sun, H., H. X. Liu, H. Xiao, R. R. He, and B. Ran. Use of Local Linear Regression Model for Short-Term Traffic Forecasting. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1836*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 143–150.
92. Zhang, X., and J. A. Rice. Short-Term Travel Time Prediction. *Transportation Research Part C*, Vol. 11, 2003, pp. 187–210.
93. Zhong, M., S. Sharma, and P. Lingras. Refining Genetically Designed Models for Improved Traffic Prediction on Rural Roads. *Transportation Planning and Technology*, Vol. 28, No. 3, 2005, pp. 213–236.
94. Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. *ASME Basic Engineering Journal*, 1960.
95. Ben-Akiva, M., G. Cantarella, E. Cascetta, J. d. Ruiter, J. Whittaker, and E. Kroes. Real-Time Prediction of Traffic Congestion. *Vehicle Navigation & Information Systems: Conference Record of Papers*, 1992.
96. Chen, H., and S. Grant-Muller. Use of Sequential Learning for Short-Term Traffic Flow Forecasting. *Transportation Research Part C*, Vol. 9, 2001, pp. 319–336.

97. Chien, S. I. J., X. Liu, and K. Ozbay. Predicting Travel Times for the South Jersey Real-Time Motorist Information System. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1855,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 32–40.

98. Chien, S. I.-J. and Kuchipudi, C. M. Dynamic Travel Time Prediction with Real-Time and Historic Data. *Journal of Transportation Engineering*, Vol. 129, No. 6, 2003, pp. 608–616.

99. Kuchipudi, C. M., and S. I. J. Chien. Development of a Hybrid Model for Dynamic Travel-Time Prediction. In *Transportation Research Record: Journal of the Transportation Research Board, No.1855*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 22–31.

100. Whittaker, J., S. Garside, and K. Lindveld. Tracking and Predicting a Network Traffic Process. *International Journal of Forecasting*, Vol. 13, 1997, pp. 51–61.

101. Nakata, T., and J.-I. Takeuchi. Mining Traffic Data from Probe-Car System for Travel Time Prediction. *Proc., 10th ACMSIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004.

102. Stathopoulos, A., and M. G. Karlaftis. A Multivariate State Space Approach for Urban Traffic Flow Modeling and Prediction. *Transportation Research Part C*, Vol. 11, 2003, pp. 121–135.

103. INRETS. Available at http://www.inrets.fr/ur/gretia/fiches/themes/Prevision_Ang_MDP_2_.htm. Accessed January 3, 2007.

104. Al-Deek, H., S. Ishak, and M. Wang. A New Short-Term Traffic Prediction and Incident Detection System on I-4. Florida Department of Transportation, University of Central Florida, 2001.

105. D'Angelo, M. P., H. M. Al-Deek, and M. C. Wang. Travel-Time Prediction for Freeway Corridors. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1676*, TRB, National Research Council, 1999, pp. 184–191.

106. Ishak, S., and H. Al-Deek. Performance Evaluation of Short-Term Time-Series Traffic Prediction Model. *Journal of Transportation Engineering*, Vol. 128, No. 6, 2002, pp. 490–498.

107. Alecsandru, C.-D. A Hyrbid Model-Based and Memory-Based Short-Term Traffic Prediction System. Louisiana State University and Agricultural and Mechanical College, 2003.

108. Dougherty, M. S., and M. R. Cobbett. Short-Term Inter-Urban Traffic Forecasts Using Neural Networks. *International Journal of Forecasting*, Vol. 13, 1997, pp. 21–31.

109. Huang, S.-H., and B. Ran. Application of Neural Network on Traffic Speed Prediction Under Adverse Weather Conditions. Presented at 82nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2003.

110. Ishak, S., and C. Alecsandru. Optimizing Traffic Prediction Performance of Neural Networks Under Various Topological, Input, and Traffic Condition Settings. *Journal of Transportation Engineering*, Vol. 130, No. 4, 2004, pp. 452–465.

111. Kisgyorgy, L., and L. R. Rilett. Travel Time Prediction by Advanced Neural Network. *Periodica Polytechnica Ser. Civil Engineering*, Vol. 46, No. 1, 2002, pp. 15–32.

112. Ledoux, C. An Urban Traffic Flow Model Integrating Neural Networks. *Transportation Research Part C*, Vol. 5, No. 5, 1997, pp. 287–300.

113. Mark, C. D., A. W. Sadek, and D. Rizzo. Predicting Experienced Travel Time with Neural Networks: A PARAMICS Simulation Study. *Proc., 7th International IEEE Conference on Intelligent Transportation Systems*, 2004.

114. Taylor, C., and D. Meldrum. Freeway Traffic Data Prediction Using Neural Networks. *Proc., 6th VNIS Conference*, 1995.

115. Wu, B., W.-J. Zhou, and W.-D. Zhang. The Applications of Data Mining Technologies in Dynamic Traffic Prediction. *Proc., IEEE Conference on Intelligent Transportation Systems*, 2003.

116. Zhang, H. M. Recursive Prediction of Traffic Conditions with Neural Network Models. *Journal of Transportation Engineering*, Vol. 126, No. 6, 2000, pp. 472–481.

117. Rumelhart, D., G. Hinton, and R. Williams. Learning Internal Representations by Error Propagation. In *Parallel Distributed Processing,* MIT Press, Cambridge, Mass.,1986.

118. Innamaa, S. Short-Term Prediction of Travel Time Using Neural Networks on an Interurban Highway. *Transportation*, Vol. 32, 2005, pp. 649–669.
119. Jiang, G., and R. Zhang. Travel-Time Prediction for Urban Arterial Road: A Case on China. *Proc., Vehicle Electronics Conference*, 2001.
120. van Lint, J. C. Reliable Real-Time Framework for Short-Term Freeway Travel Time Prediction. *Journal of Transportation Engineering*, Vol. 132, No. 12, 2006, pp. 921–932.
121. van Lint, J. W. C., S. P. Hoogendoorn, and H. J. van Zuylen. Accurate Freeway Travel Time Prediction with State-Space Neural Networks Under Missing Data. *Transportation Research Part C*, Vol. 13, No. 5–6, 2005, pp. 347–369.
122. Abdulhai, B., H. Porwal, and W. Recker. Short-Term Freeway Traffic Flow Prediction Using Genetically Optimized Time-Delay-Based Neural Networks. Presented at 78th Annual Meeting of the Transportation Research Board, Washington, D.C., 1999.
123. Annunziato, M., I. Bertini, A. Pannicelli, and S. Pizzuti. Evolutionary Feed-Forward Neural Networks for Traffic Prediction. *EUROGEN2003*, 2003.
124. Annunziato, M., and S. Pizzuti. A Smart-Adaptive-System Based on Evolutionary Computation and Neural Networks for the On-Line Short-Term Urban Traffic Prediction. *EUNITE2004*, 2004.
125. Lingras, P., and P. Mountford. Time Delay Neural Networks Designed Using Genetic Algorithms for Short Term Inter-City Traffic Forecasting. *Proc., 14th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, 2001.
126. van Hinsbergen, C. P. I., J. W. C. van Lint, and H. J. van Zuylen. Bayesian Training and Committees of State-Space Neural Networks for Online Travel Time Prediction. In *Transportation Research Record: Journal of the Transportation Research Board, No.2105*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 118–126.
127. van Hinsbergen, C. P. I., J. W. C. van Lint, and H. J. van Zuylen. Bayesian Committee of Neural Networks to Predict Travel Times with Confidence Intervals. *Transportation Research Part C*, Vol. 17, No. 5, 2009, pp. 498–509.
128. Zheng, W., D.-H. Lee, and Q. Shi. Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach. *Journal of Transportation Engineering*, Vol. 132, No. 2, 2006, pp. 114–121.
129. Zhang, H., S. G. Ritchie, and Z.-P. Lo. Macroscopic Modeling of Freeway Traffic Using an Artificial Neural Network. In *Transportation Research Record 1588*, TRB, National Research Council, Washington, D.C., 2000, pp. 110–119.
130. Amin, S. M., E. Y. Rodin, A.-P. Liu, K. Rink, and A. Carcia-Ortiz. Traffic Prediction and Management Via RBF Neural Nets and Semantic Control. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 13, 1998, pp. 315–327.
131. Huisken, G. Soft-Computing Techniques Applied to Short-Term Traffic Flow Forecasting. *Systems Analysis Modelling Simulation*, Vol. 43, No. 2, 2003, pp. 165–173.
132. Xie, Y., and Y. Zhang. A Wavelet Network Model for Short-Term Traffic Volume Forecasting. *Journal of Intelligent Transportation Systems*, Vol. 10, No. 3, 2006, pp. 141–150.
133. Coufal, D., and E. Turunen. Short Term Prediction of Highway Travel Time Using Data Mining and Neuro-Fuzzy Methods. *Neural Network World*, Vol. 3–4, 2004, pp. 221–231.
134. Xiao, H., H. Sun, and B. Ran. The Special Factor Adjustment Model Using Fuzzy-Neural Network in Traffic Prediction. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2004.
135. Yin, H., S. C. Wong, J. Xu, and C. K. Wong. Urban Traffic Flow Prediction Using a Fuzzy-Neural Approach. *Transportation Research Part C*, Vol. 10, 2002, pp. 85–98.
136. Dharia, A., and H. Adeli. Neural Network Model for Rapid Forecasting of Freeway Link Travel Time. *Engineering Applications of Artificial Intelligence*, Vol. 16, 2003, pp. 607–613.
137. Jiang, X., and H. Adeli. Dynamic Wavelet Neural Network Model for Traffic Flow Forecasting. *Journal of Transportation Engineering*, Vol. 131, No. 10, 2005, pp. 771–779.

138. Li, S. Nonlinear Combination of Travel-Time Prediction Model Based on Wavelet Network. Presented at the IEEE 5th Conference on Intelligent Transportation Systems, 2002.

139. Tan, G., W. Yuan, and H. Ding. Traffic Flow Prediction Based on Generalized Neural Network. Proc., IEEE Conference on Intelligent Transportation Systems, 2004.

140. Ishak, S., P. Kotha, and C. Alecsandru. Optimization of Dynamic Neural Network Performance for Short-Term Traffic Prediction. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1836*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 45–56.

141. Yasdi, R. Prediction of Road Traffic Using a Neural Network Approach. *Neural Computing and Applications*, Vol. 8, 1999, pp. 135–142.

142. van Lint, J. W. C., S. P. Hoogendoorn, and H. J. van Zuylen. Freeway Travel Time Prediction with State–Space Neural Networks: Modeling State–Space Dynamics with Recurrent Neural Networks. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1811*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 30–39.

143. Yun, S.-Y., S. Namkoong, J.-H. Rho, S.-W. Shin, and J.-U. Choi. A Performance Evaluation of Neural Network Models in Traffic Volume Forecasting. *Mathematical and Computer Modelling*, Vol. 27, No. 9–11, 1998, pp. 293–310.

144. Dia, H. An Object-Oriented Neural Network Approach to Short-Term Traffic Forecasting. *European Journal of Operational Research*, Vol. 131, 2001, pp. 253–261.

145. Bajwa, S. u. I. Short-Term Travel Time Prediction Using Traffic Detector Data. University of Tokyo, 2003.

146. Bajwa, S. u. I., E. Chung, and M. Kuwahara. An Adaptive Travel Time Prediction Model Based on Pattern Matching. Proc., 11th Intelligent Transport Systems World Congress, 2004.

147. Clark, S. Traffic Prediction Using Multivariate Nonparametric Regression. *Journal of Transportation Engineering*, Vol. 129, No. 2, 2003, pp. 161–168.

148. Handley, S., P. Langley, and F. A. Rauscher. Learning to Predict the Duration of an Automobile Trip. Proc., Fourth International Conference on Kownledge Discovery and Data Mining, 1998.

149. Kim, T., H. Kim, and D. J. Lovell. Traffic Flow Forecasting: Overcoming Memoryless Property in Nearest Neighbor Nonparametric Regression. Proc., 8th International IEEE Conference on Intelligent Transportation Systems, 2005.

150. Ohba, Y.,H. Ueno, and M. Kuwahara. Travel Time Prediction Method for Expressway Using Toll Collection System Data. Proc., 7th World Congress on Intelligent Systems, 2000.

151. Oswald, R. K., W. T. Sherer, and B. L. Smith. Traffic Flow Forecasting Using Approximate Nearest Neighbor Nonparametric Regression. Center for Transportation Studies, University of Virginia, 2001.

152. Otokita, T., K. Hashiba, and T. Oda. Travel Time Prediction Based on Pattern Extraction from Database. Proc., 5th World Congress on Intelligent Transportation Systems, 1998.

153. Smith, B. L., and M. J. Demetsky. Multiple-Interval Freeway Traffic Flow Forecasting. In *Transportation Research Record 1554*, TRB, National Research Council, Washington, D.C., 1996, pp. 136–141.

154. You, J., and T. J. Kim. Development and Evaluation of a Hybrid Travel Time Forecasting Model. *Transportation Research Part C*, Vol. 8, 2000, pp. 231–256.

155. Li, L., W.-H. Lin, and H. Liu. Type-2 Fuzzy Logic Approach for Short-Term Traffic Forecasting. Proc., IEEE Conference on Intelligent Transport Systems, 2006.

156. INRIX. Available at http://www.inrix.com. Accessed Jan. 3, 2007.

157. Smola, A. J., and B. Scholkopf. A Tutorial on Support Vector Regression. *Statistics and Computing*, Vol. 14, 2004.

Neural Networks for Travel Time Prediction on Interrupted Flow Facilities

GHASSAN ABU-LEBDEH
American University of Sharjah

The previous article has provided a broad overview of artificial intelligence (AI) and traffic prediction. In this article, the focus is on a specific problem (namely the prediction of route travel time on arterials) and a specific AI paradigm (neural networks). This article can thus be regarded as a specific case study or an illustrative example for what was discussed in the previous article.

Travel time is an easily perceptible measure of traffic conditions on arterials or on freeways. However, there has been limited research on travel time estimation and prediction on major interrupted flow facilities as signalized arterials compared to continuous (uninterrupted) flow facilities as freeways. The vastness of arterial systems, complexity of traffic behavior on arterials especially at signalized intersections, and lack of traffic surveillance coverage compared to freeways are among the probable reasons. This implies that travel time prediction on urban arterials, at least in the near term, should rely only on existing technologies, be it computing, detection, or sensing. Easily available data that can be obtained in the field using already-in-use traffic detection technologies like loop detectors or other image processing-based technologies would be the primary source since a major investment in detection equipments for arterials is not likely in the short term, although initiatives such as vehicle infrastructure integration (VII) or the Connected Vehicle research program may change that once implemented. Existing and recently matured advanced computation paradigms, such as neural networks (NNets), enable us to do more with fewer resources where we can squeeze far more knowledge and insights from ordinary data.

This article examines the use of NNets for travel time estimation and prediction on arterials; it thus serves as an in-depth case study for a specific application among the several approaches described in the previous overview article by van Lint and van Hinsbergen. It is shown how the state–space notion of roadway traffic can be used with NNets to model travel time on urban arterials. The state–space representation of complex dynamical systems such as arterials provides enough insight to model it accurately. State–space NNets (SSNN) is a generic form of recurrent neural networks (RNN), and is particularly suited for the purpose of travel time predictions on arterials. Furthermore, when SSNN are combined with conditional independence (CI) graphs, a specialized statistical technique, to analyze the independence and interaction among variables involved in the travel time process, a more efficient and robust travel time prediction process emerges whereby fewer variables are needed. This modeling approach relies on data that is easily available in the field and hence has great potential in advanced traveler information systems (ATIS) and advanced traffic management systems (ATMS) applications.

Although the use of NNets is presented here in the context of a transportation problem, the approach described applies to any system whereby values of key system performance parameters are to be predicted at a future point based on a current state of a system and its immediate past.

BACKGROUND

Travel time prediction on arterial networks is still lagging behind as compared to that of freeways. Vlahogianni et al. gives a critical discussion of the short-term traffic forecasting techniques for freeways and urban arterials for different types of implementations including ATIS and ATMS (*1*). This review asserts that AI-based nonparametric modeling techniques like artificial NNets (ANNs), as was discussed in the previous article by van Lint, is promising for traffic forecasting problems as it gives robust and accurate models. Travel time prediction models have been proposed by various researchers for freeways (*2–4*), and there is significant volume of studies related to the application of different types of ANNs in short-term travel time prediction on continuous-flow facilities as freeways (*5–7*), however, travel time prediction on interrupted-flow facilities as urban arterials is still a fairly new domain. The work by Singh describes new models based on SSNN that performed well in both congested and uncongested conditions (*8*).

Sisiopiku and Rouphail presented a literature review of methods for travel time estimation on urban arterials and concluded that most models are link specific as opposed to path level (*9*). Stathopoulos and Karlaftis proposed a multivariate time series state–space model for traffic flow prediction (*10*). The model of Lin et al. relies on existing intersection delay formulas, which are not suitable for over-saturated conditions (*11*). The model of Liu et al. is for a particular stretch of urban arterials and does not consider geometrics of arterials as one of the variables nor does it explicitly consider signal control settings as independent variables (*12*). A new study by Saadaoui used the SSNN approach to study travel time in the presence of random incidents (*13*). The recent work of Zheng and van Zuylen on the subject of urban travel time estimation is interesting and unique (*14–16*). While earlier studies were all about predicting a value for travel time (by implication, a mean value), Zheng and van Zuylen developed a theoretical travel time distribution that can capture the dynamics of travel time. The proposed distribution of travel times explicitly considers the uncertainty in delay at signalized intersection as well as the inherent uncertainty in other parameter (e.g., free-flow travel time). Model results were compared with VISSIM and field data. Based on observed field data, the proposed models could represent field travel time distribution reasonably well. Also, the travel time distributes derived from the model matched those from VISSIM for different congestion scenarios. When corridor travel times were considered (with explicit consideration of signal coordination but not platoon dispersion), results from the proposed models compared favorably with observed and simulated travel times. Besides adeptly formulating models to capture the dynamics of important components of the urban travel time components, the results of this work bring into focus the importance and practical implications of considering (or ignoring) the inherent variability in travel time. Though these recent studies on the topic provide reasonable methodologies for travel time prediction, a generalized model that captures turning movements, geometrics of arterials, and signal control settings is still lacking. Specifically, there is need for a travel time estimation and prediction model or an approach that relies on traffic variables that are easily obtainable in the field. Inclusion of geometrics and signal control is necessary to provide generalized models applicable to real world arterial networks. The combination of the CI graphs and SSNN provides a generalized approach to modeling travel time on real-world arterial networks.

THE PROBLEM OF TRAVEL TIME PREDICTION

Although they have many similarities, there are slight differences between travel time prediction and estimation. The focus here is on short-term prediction of travel time. However, most of the technicalities described apply to estimation as well. For now, estimation of travel time is an intermediate step of travel time prediction, and for the purpose of this presentation it is assumed to be available as input to the travel time prediction process. Travel time estimation using SSNN can found in the work described in Singh (*8*). This section describes the problem and outlines the typical objective of such problem.

First, the terminology is presented. Travel time is the average travel time; this is the average value of the travel time incurred by individual vehicles if each vehicle can be tracked on the arterial link. Traffic conditions on an arterial link at any time period can be represented by a set of traffic variables like average flow rate, average speed, queue length, etc. State of a link is the term used to describe traffic conditions on a link at a given time. This is an important input in estimating or predicting average travel time. The estimated average travel time on an arterial link is the experiential travel time by vehicles traversing that link. Experiential denotes travel time that has already been experienced by vehicles that have departed in the current or past time periods when the state of the system was known. The predicted average travel time on an arterial link is travel time that will be experienced by vehicles when their departure is in a future time period for which the state of the link is currently unknown.

The inherent distinction between estimation and prediction of average travel time should be noted in the above definitions. Travel time estimation is done when the vehicle has already departed and has experienced the travel time in its trip. The average travel time prediction is a different problem from the estimation as it requires a modeling technique to capture the dynamic (i.e., time dependent) relationship between current state and the future travel time. This presentation here is for models for predicting average travel time for a short term into the future, although the same general approach may be used for travel time estimation.

TRAVEL TIME MODELING: RELEVANT FACTORS AND PROBLEM FORMULATION

A large number of variables impact travel time of different traffic movements. These are: (*a*) traffic demand variables including average flow rate for through, left-, and right-turning movements denoted as V_T, V_L, and V_R, respectively; (*b*) average speed for through, left-, and right-turning movements denoted, respectively, as S_T, S_L, and S_R; (*c*) signal control variables including green interval for an approach (*G*), and offset between a pair of signalized intersections for an approach (Off); (*d*) queue length variables including queue length in through and right-turning movement lanes denoted as Q_T, and queue length in left-turning lane denoted as Q_L; (*e*) speed limit posted on the link (S.L); and (*f*) geometrics and geometrics-influenced attributes.

Now, the ideal travel time (ITT) for through and right-turning movement (ITT$_T$) is the summation of the ratio of link length to speed limit (S.L.) plus the green interval for the left-turning movement (since during that period, the through and right-turning movements are stopped at the signal). ITT is considered the same for through and right-turning movements as both of these movements generally share the same lanes and signal phase. The left-turn movement generally has a separate turning bay or pocket and a distinct signal phase. The ITT for

the left-turning movement is denoted by ITT_L. ITT_L is the summation of the ratio of link length to the speed limit (S.L.) plus the green interval for the through movement and the red interval for the approach. This implies a phasing scheme of exclusive left-turn phase with leading green. Other phasing schemes can be considered, in which case the definition of ITT_T and ITT_L will have to change accordingly. The variables listed above are highly interrelated. The fluctuations in average travel time resulting from the interactions among those variables make it a dynamic and complex spatial–temporal function of traffic demand and the capacity supply of the link (or links of the route) under consideration.

State–Space Notion of Traffic and Modeling of Travel Time

Let an arterial route in a network consists of n arterial links represented as $X_1, X_2, X_3, \ldots, X_n$. Let t_p^* be the current time period of departure of a vehicle that starts from an entry point (upstream intersection) of link X_1 and ends its trip by clearing the intersection at the end of link X_n. $t_p^* = [t_0^*, t_1^*]$, where t_0^* and t_1^* are the starting and ending times of the time period t_p^*. Traffic variables that shape the prevailing traffic conditions on an arterial network during a current departure time period t_p^* can be measured in the field through traffic detection system, or can be estimated through other measurable traffic variables. Take time periods to be short and assume that traffic conditions are constant within each time period. Here, the length of each time period is taken as 5 min and is called the aggregation interval Φ, thus $\Phi = t_1^* - t_0^*$ [a detailed discussion and justification for selecting 5 min as an optimal aggregation interval is provided in Singh (8)]. Let the state of link X_1 during t_p^* be represented by the vector $V(X_1, t_p^*)$ which denotes the known traffic conditions on link X_1 during current time period t_p^*. Denote the average travel time to traverse link X_1 during the current time period t_p^* by TT (X_1, t_p^*). The modeling approach to prediction first estimates the current time period average travel time on the basis of the current state of the link (since it is an input into the prediction process). The estimation process is not covered here but bears many similarities to the prediction process.

      Assume that the average travel time is to be predicted for a future departure time period which is starting just when the current departure time period t_p^* ends (i.e., at time instant t_1^*). Let Δ be the prediction horizon which is a short time interval in the future for which the average TT is to be predicted. The future time period that begins at t_0 and ends at t_1 is denoted as $t_{p+1\Delta} = [t_0, t_1]$. This implies that $t_0 = t_1^*$, and the length of the future time period $t_{p+1\Delta}$ is simply the addition of a prediction horizon length Δ to the ending time of the current time period t_p^* (i.e., t_1^*). So, $t_{p+1\Delta} = t_1^* + \Delta$. Since the future time period of $t_{p+1\Delta}$ is a single multiple of the prediction horizon length Δ, it is called a one-step future time period, and the predicted value of the average TT for this one-step future time period is called the one-step predicted average TT. Conveniently, the prediction horizon Δ can be assumed to be the same as the aggregation interval Φ. So, $\Delta = \Phi$ (e.g., 5 min). This assumption is made so that the temporal scale at which the state of a link and the TT are analyzed can be the same for all time periods, be it current, past, or future. Moreover, it makes the task of updating the state of a link and the TT values that are evolving with time periods easier. The predicted average TT to traverse link X_1 during the one-step future time period $t_{p+1\Delta}$ is then denoted by TT $(X_1, t_{p+1\Delta})$. The problem of one-step average TT prediction can now be defined as follows: the one-step average TT prediction on an arterial link X_1 is the modeling problem to approximate the underlying function between the one-step predicted average TT [TT $(X_1, t_{p+1\Delta})$] during the one-step future time period $t_{p+1\Delta}$ on one hand, and the state

$[V(X_1, t_p^*)]$ and the estimated average TT $[X_1, t_p^*)]$ on link X_1 during the current time period t_p^* on the other.

CI Graph for Modeling Travel Time

CI graphs are the founding stones of a series of probabilistic graphical models like the Markov networks and the Bayesian belief networks (*17*). They are used to study the interaction between dependent and independent variables. Here, the CI graphs are introduced as relevant computational tools that can be used to analyze the interaction and conditional independence between TT and the traffic variables that represent the state of a link and thus augment the use of SSNN. The CI graph and the Kullback–Leibler information divergence are used here for two reasons: (*a*) understand TT estimation and prediction; the Kullback–Leibler information divergence is helpful in evaluating the amount (quantity) of information contained in traffic variables about TT hence increasing the understanding of this process, and (*b*) select the parent input variables: the use of the Markovian property of the CI graph identifies a set of variables termed "parent" input variables. These parent input variables give sufficient information about TT that the remaining variables are redundant (i.e., can be dropped from the input vector). In addition, the CI graph and the Kulback–Leibler information are used here as a mean to improve the efficiency of the SSNN (described later). First, construction of a CI graph for average TT estimation is illustrated (for prediction, the only difference is that the estimated TT_T of the previous time period becomes another input).

Assume that the distribution of TT and the variables that shape the state of the link is multivariate normal. Hence all variables can be represented in the sample correlation matrix, which depicts the degree of linear association between TT and the independent variables shaping the state of the link. But this correlation matrix also shows that the independent variables are correlated among themselves. Hence, this matrix does not provide useful information if the dependence of the TT is to be analyzed taking into account intercorrelations among independent variables; a scaled inverse of the correlation matrix is more useful. It is noted here that the assumption that TT and the other traffic variables being normally distributed holds well when a large sample space is considered. Since data collection from well-equipped real-world networks, or from microsimulation models if synthetic data are to be used, provides a large sample space, the traffic variables' joint distribution can be considered as normal.

The inverse matrix of the sample correlation matrix is calculated and scaled such that the diagonal elements of the inverse matrix equal to 1, as shown in Table 1. This gives useful information, which is then used for constructing the CI graph. The off-diagonal elements of the scaled sample inverse correlation matrix are the negatives of the partial correlation coefficients between the corresponding pairs of variables given the remaining variables. If the off-diagonal elements are zero then the corresponding variables are conditionally independent given the remaining variables. This scaled inverse correlation matrix is approximated such that the elements that have values close to zero (here <0.1) are considered zero, and values ≥0.1 are made nonzero entities.

For the CI graph $G = (K, E)$, K represents the nodes and E represents the edges between these nodes. Each variable in the TT modeling process is considered as a node, thus for this section $K = 14$ (since there are 13 independent variables affecting travel time, and the 14th variable is the travel time itself as discussed above). An edge is put between each two variables if there is a nonzero entity (≥0.1; see Table 1); no edge otherwise. The resulting graph is the CI

**TABLE 1 Sample Scaled Inverse Correlation Matrix for Average TT,
Through Movement TT Estimation***

	V_T	V_L	V_R	G	Off	S.L.	ITT_T	ITT_L	Q_T	Q_L	S_T	S_L	S_R	TT_T
V_T	1.0													
V_L	–0.7	1.0												
V_R	–0.3	–0.4	1.0											
G	–0.1	–0.1	–0.2	1.0										
Off	0.0	–0.0	0.0	0.1	1.0									
S.L.	–0.0	0.2	0.2	0.0	–0.2	1.0								
ITT_T	–0.0	–0.0	–0.1	0.3	0.2	–0.6	1.0							
ITT_L	0.0	0.0	0.0	0.0	0.0	–0.0	–0.9	1.0						
Q_T	–0.2	–0.2	–0.4	–0.0	0.1	0.0	0.0	–0.2	1.0					
Q_L	0.0	–0.3	–0.2	–0.1	–0.1	0.0	–0.0	0.2	–0.6	1.0				
S_T	0.0	–0.1	0.0	–0.2	–0.1	–0.2	–0.0	–0.7	0.1	0.3	1.0			
S_L	–0.0	0.2	0.0	–0.0	0.1	–0.1	0.0	–0.1	–0.3	0.0	–0.8	1.0		
S_R	0.0	–0.2	–0.0	0.0	–0.1	0.0	–0.0	0.1	0.3	0.0	–0.3	–0.8	1.0	
TT_T	0.0	0.0	0.0	0.1	–0.0	0.1	0.0	–0.4	–0.4	–0.3	0.2	–0.3	–0.4	1.0

*For prediction, a similar table would be produced, but TT_T for current time period would be included as an additional variable.

graph for the average TT modeling process (Figure 1). There are two types of edges: directed and undirected. The directed edge is drawn if the general understanding of the process indicates that the variable at the head of the directed edge is influenced by the variable at the tail. For example, TT_T is influenced by all other variables as they represent the state of the link. Similarly, if the general understanding of the process is not able to detect causation between variables, the edge is an undirected edge (shown as dashed line).

The CI graph can be interpreted by the global Markovian property which implies that any two subsets of variables separated by a third are independent conditionally only on variables in the third subset (*14*). Using this global Markovian property, the so-called parent input variables are selected. The parent input variables are a set of variables obtained from the CI graph which contain sufficient information that the remaining variables do not add any extra information (to travel time estimation–prediction).

Figure 1 shows that the travel time of the through movement (TT_T) is directly interacting with eight variables: Q_T, Q_L, S_T, S_L, S_R, G, S.L., and ITT_L. Following the global Markovian property, this can be written as $\{TT_T \perp\!\!\!\perp (V_T, V_L, V_R, \text{Off}, ITT_T) \mid (Q_T, Q_L, S_T, S_L, S_R, G, \text{S.L.}, ITT_L)\}$, where the $\perp\!\!\!\perp$ sign partitions the sets of variables that are conditionally independent, and the $\mid$ sign partitions the conditional independence just stated and the set of variables which makes that possible (i.e., the set of variables which, if known, the first two sets of variables are conditionally independent). The set of variables Q_L, S_T, S_L, S_R, G, S.L., and ITT_L is the parent input variables. By definition of CI, the parent input variables are sufficient to model the average TT.

After identifying the parent input variables, the concept of Kullback–Leibler information divergence and information proper (*17*) can be used to evaluate the "quantity" of information contained in each of these variables about TT_T. The information proper is shown in Figure 1 in parentheses along with the directed edges that are connecting the parent input variables to TT_T. Here, the parent input variables are used as input variables to the SSNN for average TT prediction for through movements.

FIGURE 1 CI graph for average TT, through movement estimation (for prediction, similar graph would be produced but TT$_T$ for current time period would be included as an additional variable).
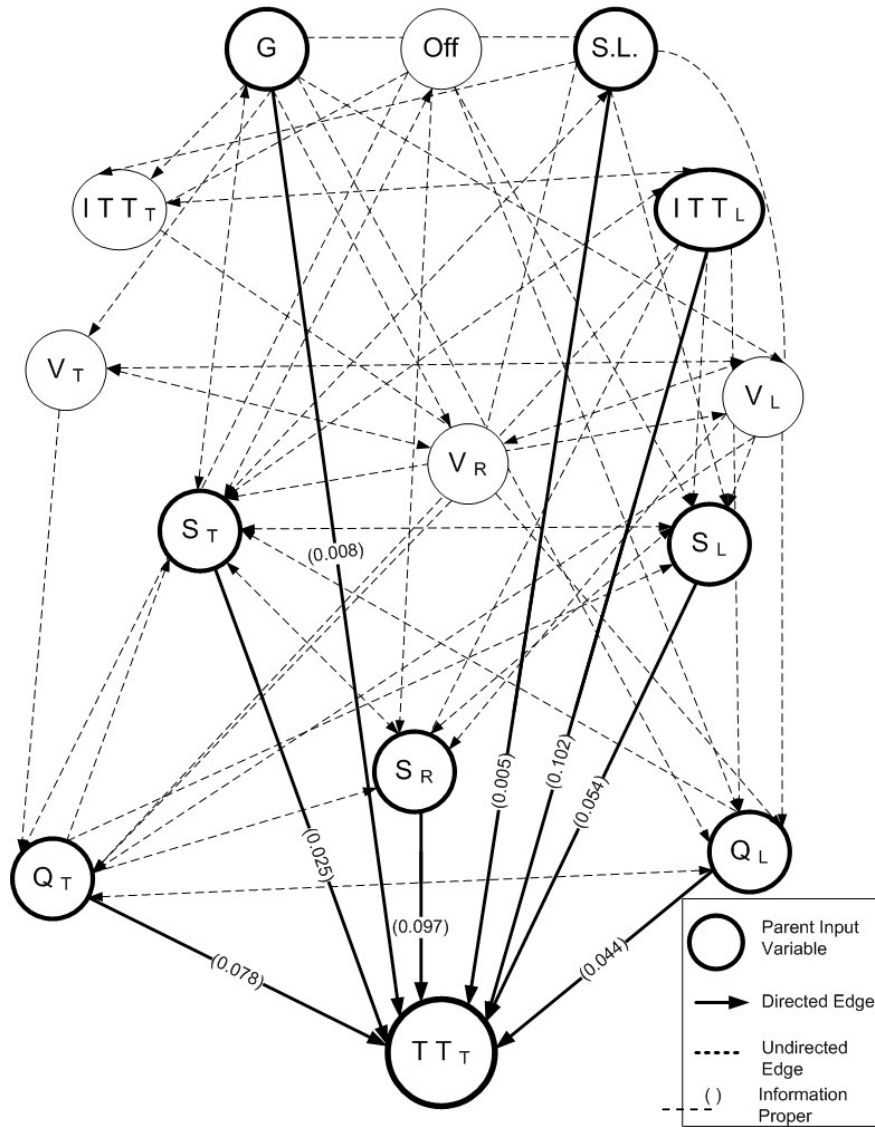

SSNN: A Brief Description

The SSNN is a generic form of the RNN with a state–space representation of the process being modeled. The state–space representation of a SSNN is mathematically stated as state–space equations. Let a nonlinear time-variant dynamic system be noise free (*18*), then,

$$x(t+1) = \varphi\ (W_a x(t) + W_b u(t))$$
$$y(t) = C\ x(t)$$

where

q = number of unit delays used to feed the output of the hidden layer back to the input layer (order of the model);

m = number of input variables;

p = number of output variables;

t = discrete time point;

$x(t)$ = q-by-1 vector denoting the state of a nonlinear discrete time system;

$u(t)$ = m-by-1 vector denoting the input applied to the system;

$y(t)$ = p-by-1 vector denoting the corresponding output of the system;

W_a = q-by-q matrix;

W_b = q-by-$(m+1)$ matrix;

C = p-by-q matrix; and

φ = a nonlinear mapping function.

The above state–space equations suppose that at each discrete time point t, a vector valued observation $y(t)$ is related to $x(t)$. The $x(t)$ is termed state of the system at discrete time point t. The second equation representing the relationship between $y(t)$ and $x(t)$ is called the observation equation. The observation equation is shown as a linear function of the state of the system, but could also be a nonlinear function. The first equation is termed a transition or state equation, as it models the state of the system at any time point t. The state of the system at any time point $t+1$ depends on the state at the previous time point t and the external inputs to the system $u(t)$. The nonlinear mapping function φ captures the relationship between the state at the current time point on one hand and the state in the past time point and the external input vector on the other. An assumption made in the above state–space equations is that the system is noise free, which is hard to achieve when traffic data are collected using detection technologies. In this presentation that point is not addressed. For real-world implementations, noise in data must be accounted for.

SSNN Topology for TT Prediction

The SSNN model for TT prediction is shown in Figure 2. The figure shows an input layer which has a vector of input variables [$U(t)$] at the current time period t that affect TT. This input layer is connected to the hidden layer by a trainable connection where the connection weights are trained by a gradient descent algorithm. The hidden layer in the network represents the state of the system at time period $t+1$. This state at time period $t+1$ activates the nodes in the context layer which represent the state of the system in the current time period t. The context layer has a trainable connection with the hidden layer which provides information about the state of the current time period to the hidden layer. So, the state of the system (here, an arterial link) at a future and a current time period are captured in the internal hidden and context layers, respectively. The hidden layer is then connected to the output layer by a trainable connection. The output layer gives the average TT in the future time period. Hence, this SSNN topology is for average TT prediction knowing the current state traffic variables and the current estimated average TT.

The effectiveness of SSNN may decrease because of the complexity in the topology which in turn appears because of the inclusion of more variables than necessary. So, the CI graph was combined to reduce the number of variables needed as inputs to model the average travel
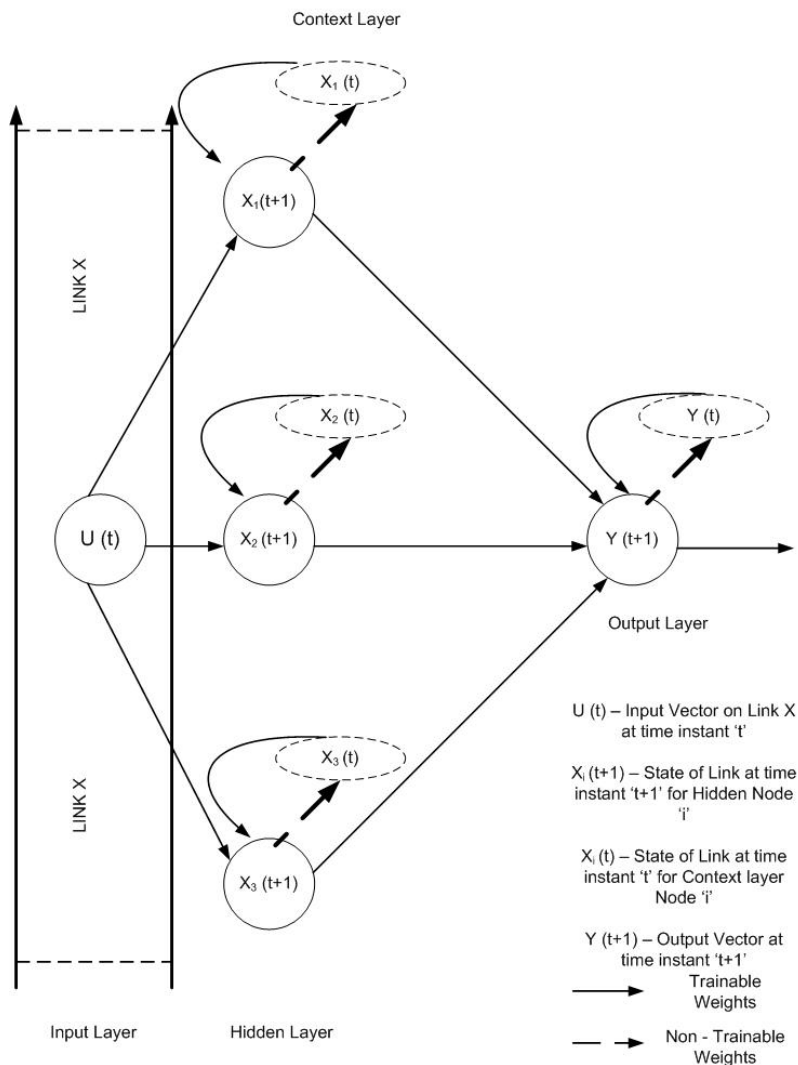
**FIGURE 2 SSNN topology adopted for TT prediction on an arterial link
for a specific traffic movement**.

time. The combined models are labeled CI–SSNN. All equal, the results (presented later) showed that that the CI–SSNN models are superior to the SSNN models.

EXPERIMENTAL SET UP AND DATA

Synthetic traffic data are in this experimental set up. It is much easier to obtain, and more of it can be generated exactly as necessary. This is done strictly for convenience; however, more importantly, the use of such data is incidental to the objective of this presentation and does not diminish the value of the modeling approach (SSNN). If an SSNN can work and be useful with synthetic data it would be just as useful with real-world data of comparable quality. TSIS-CORSIM, a traffic microscopic simulation model, was used in conjunction with a hypothetical

urban arterial network to create synthetic TT data for various geometric and demand combinations using the 25 signalized intersection network shown in Figure 3. A representative sample of four arterials with various traffic demand, geometry, and signal controls was selected from the network as a source of data (arterials 6-10, 16-20, 21-25, and 3-23). All selected links have two lanes in each direction and each intersection approach has a left-turn pocket. Speed limits were 30 mph for short links (1,500 ft) and 45 mph for the rest.

Nineteen 5-min periods (5,700 simulation seconds) were used whereby traffic demands were varied widely (no congestion then gradually to congested and then gradually back to no congestion; see Figure 4). Conditions were assumed constant for each 5-min period. Signal control was independently optimized based on demand. In dynamic signal control systems, signal timings are optimized based on the temporal variations in the estimated or predicted TT. Since this study takes signal timing parameters as independent variables for the purpose of TT prediction, the optimized signal timing after each cycle or in a prespecified time period, say
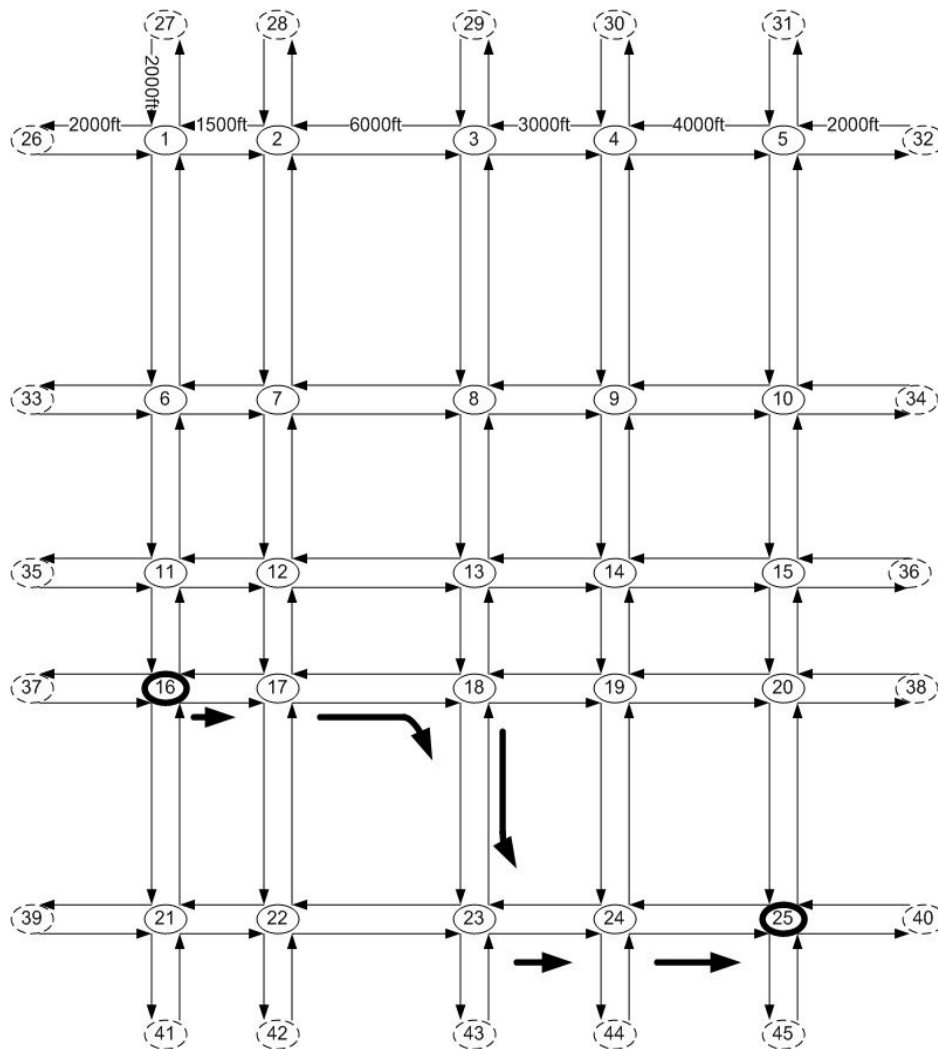
FIGURE 3 Urban arterial network designed in TSIS-CORSIM for TT estimation and prediction. Example route from Intersection 16 to Intersection 25 is shown.

5 min as in this study, can, alternatively, be considered instead of static signal timing for each intersection. Turning movements were explicitly considered. Left- and right-turn percentages were kept at 20% and 10%, respectively. The phasing scheme is four phases with exclusive leading left-turn phase. The average speed (S_T) was first estimated and validated using regression from easily observable traffic variables like V_T, V_L, V_R, G, Off, Q_T, Q_L, S.L., ITT_T, and ITT_L. This average speed is for the entire link; it is not speed at a point on a link.

RESULTS AND DISCUSSION

CI-SSNN models were trained and tested to evaluate their performance and efficiency to model travel time. During training the input patterns were presented sequentially rather than as concurrent matrix. The learning rate and momentum values were 0.0001 and 0.9, respectively. Four hidden and context layer nodes were found to produce the best results. First, the SSNN performance for TT estimation was compared with and without the CI graphs method. It is clear from the results in Table 2 that the CI-SSNN model performance is superior. The remainder of the results is based on the CI-SSNN models. The correlation coefficient (R^2), mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) were used as measures of effectiveness. The results shown in Table 3 are for the CI-SSNN models for predicting TT of through traffic (CI-SSNN-Thru). The results shown are for the testing set.

Table 3 shows that the CI-SSNN-Thru prediction model performs very well on the testing set. This ensures that the proposed model is generalized and can predict TT for new traffic patterns presented to the model but not used in training.

Figure 5 shows the TT patterns in the testing set for arterial 6-10 which has four links, namely 6-7, 7-8, 8-9, and 9-10. For each link on the arterial the observed and modeled TT are plotted against the departure times during the study period (minute 0 to minute 95 in 5-min increments). Comparable results were obtained for other arterials.
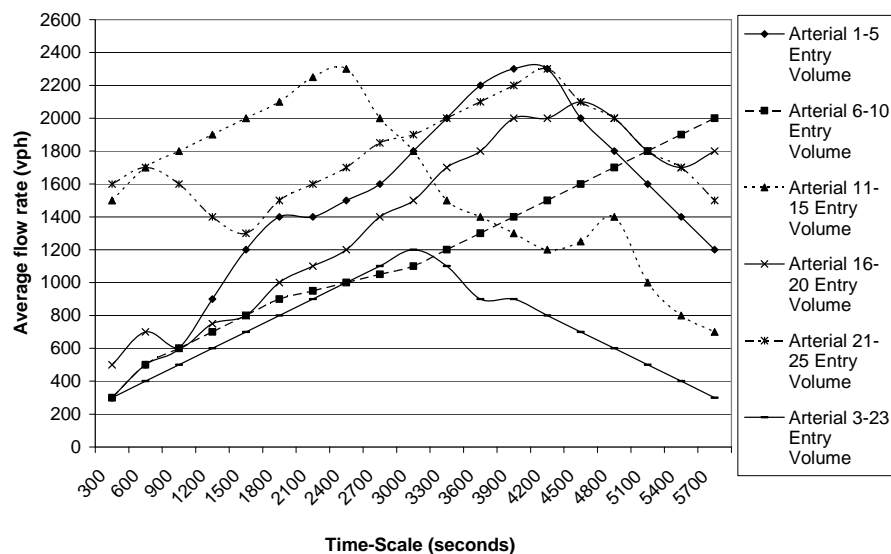
FIGURE 4 Average flow rate and variations at entry nodes of network arterials.

**TABLE 2 Performance Measures of CI-SSNN-Thru and SSNN-Thru
Estimation Model on Testing Set***

Arterial	MAE, s	RMSE, s	MAPE, %
6-10	7.3 (9.3)	10.3 (11.9)	7.7 (11.0)
16-20	5.6 (6.6)	7.0 (8.7)	6.5 (8.4)
21-25	6.2 (4.7)	9.0 (6.3)	6.5 (5.2)
3-23	7.1 (7.0)	9.0 (9.0)	8.5 (8.5)
Total testing set	6.6 (6.9)	8.9 (9.2)	7.3 (8.2)

*Values in parentheses are for SSNN.

Based on the results in Figure 5, the following two inferences can be made about the CI-SSNN-Thru model: (*a*) the predicted average TT lies in the same range as the actual (observed) average TT for all links of the arterial. This confirms that the CI-SSNN-Thru model is able to both learn the process and generalize it, and hence it can be used to predict average TT within the range that is typical of a particular link, and (*b*) close inspection of the TT patterns of each arterial link shows that the predicted average TT follows the general pattern of the observed values although in some cases the CI-SSNN model appears to have a slight "delay" in responding to changes in observed TT. Given the complex nature of traffic flow dynamics and the temporal nature of TT, the closeness of modeled to the observed values is satisfactory; the CI-SSNN model is able to capture the dynamic and temporal aspects of TT quite well. Similar CI-SSNN models were developed for left and right turns but not shown here. The results of CI-SSNN model for prespecified routes (that include links of different arterials and all three types of traffic turns) are as follows: (also see Figure 6). In this case the MAPE for the route TT was less than 4% (Table 4). The % MAPE was considerably lower than those of the individual links due to the averaging of errors of all three movements. Since we typically are interested in TT of a route more than a link, this outcome is favorable. A different depiction of this result is shown in Figure 6 which shows the pattern of the actual and predicted travel for the individual links that made up the same route. It is noted that the actual and predicted TT are in the same general range and that there is a consistency in the overall trend between the two.

TABLE 3 Performance Measures of CI-SSNN-Thru Prediction Model on Testing Set

Arterial	MAE, s	RMSE, s	MAPE, %
6-10	8.9	13.2	9.3
16-20	7.5	9.0	8.8
21-25	7.7	11.1	8.2
3-23	7.5	9.2	8.7
Total testing set	7.9	10.7	8.8

**TABLE 4 Performance Measures of CI-SSNN-Thru Prediction Model
on Testing Set for a Selected Route**

Route	MAE, s	RMSE, s	MAPE, %
16-17-18-23-24-25	18.4	23.2	3.8
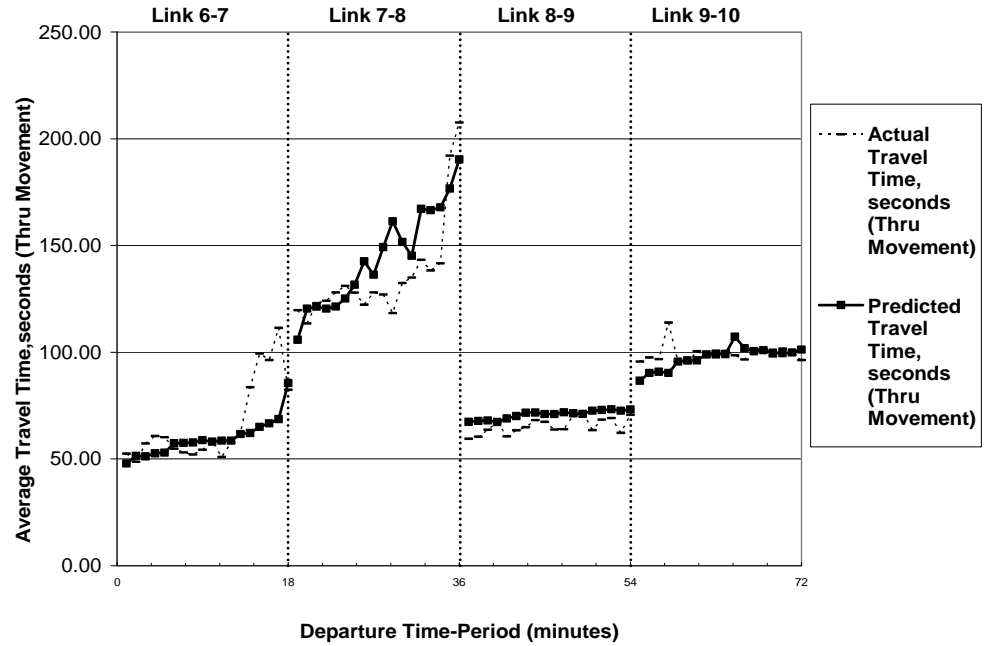
FIGURE 5 Pattern of actual and predicted TT for through movements of Arterial 10 (testing set results shown).
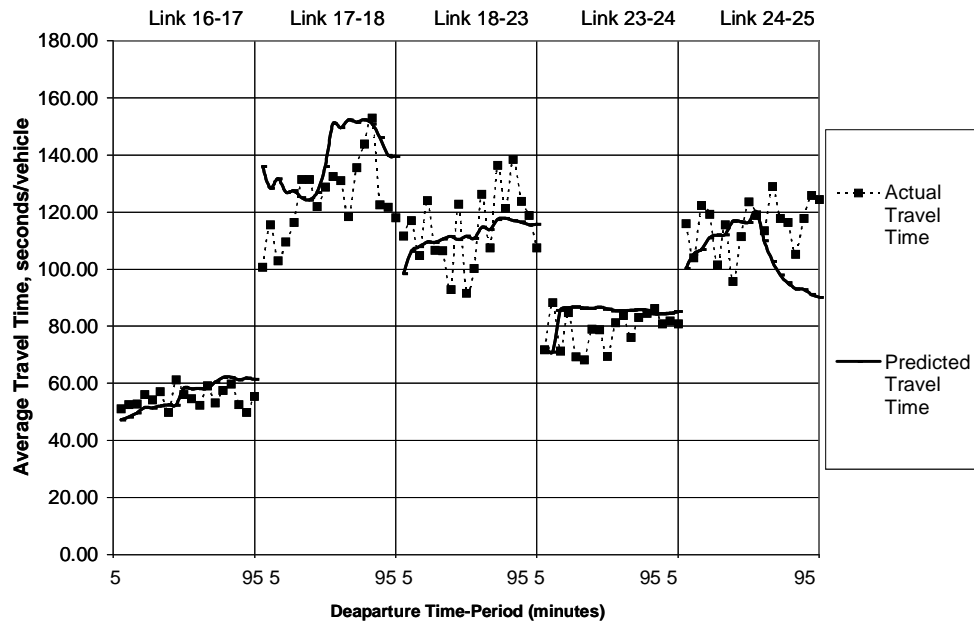
FIGURE 6 Pattern of actual and predicted TT for links of route 16-17-18-23-24-25.

CONCLUSIONS AND RECOMMENDATIONS

This article presents conditional independence graphs and state–space NNet-based (CI-SSNN) models to predict short-term TT on signalized urban arterials using easily observable traffic data including average flow rate, queue length, geometrics, and signal control parameters. The reliance of these models on easily available data demonstrates the feasibility and promise of the proposed approach for real-time applications. The state–space dynamics of traffic behavior in an urban arterial is captured successfully using SSNN since they provide flexibility and efficiency while modeling the nonlinear dynamical systems. The CI-SSNN models performed equally well on congested and uncongested traffic conditions thus demonstrating much needed robustness for successful field implementation. A preprocessing of the easily observable traffic parameters was used to minimize the number of input parameters. For this, the CI graphs technique was used to enable the analysis and assessment of the independence and interaction among the observable traffic parameters on one hand and the average TT on the other, thus increasing the understanding of the TT modeling process as well as improving the predictive performance of the SSNN. The flexibility of the modeling approach allows the inclusion of these factors if relevant data are available. The proposed approach did not account for such relevant factors as weather and incidents, and hence it is recommended that the scope of the models be expanded to include incident and weather variables so that they may be applied in the field. It is recommended that, as applicable, the CI graphs technique be coupled with ANN models to remove unnecessary variables.

REFERENCES

1. Vlahogianni, E. I., J. C. Golias, and M. G. Karlaftis. Short-Term Traffic Forecasting: Overview of Objectives and Methods. *Transport Reviews*, Vol. 24, No. 5, 2004, pp. 533–557.
2. Zwet, E., and J. Rice. A Simple and Effective Method for Predicting Travel Times on Freeways. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, No. 3, 2004, pp. 200–207.
3. Wu, C. H., J. M. Ho, and D. T. Lee. Travel-Time Prediction with Support Vector Regression. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, No. 4, 2004, pp. 276–281.
4. Chien, S. I., and C. M. Kuchipudi. Dynamic Travel Time Prediction with Real-Time and Historic Data. *Journal of Transportation Engineering*, Vol. 129, No. 6, 2003, pp. 608–616.
5. Mark, C. D., and A. W. Sadek. Learning Systems for Predicting Experiential Travel Times in the Presence of Incidents: Insights and Lessons Learned. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1879*, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 51–58.
6. van Lint, J. W. C. Reliable Travel Time Prediction for Freeways. PhD thesis. Delft University Press, Delft, Netherlands, 2004.
7. van Lint, J. W. C., S. P. Hoogendoorn, and H. J. van Zuylen. Freeway Travel Time Prediction with State–Space Neural Networks: Modeling State-Space Dynamics with Recurrent Neural Networks. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1811*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 30–39.
8. Singh, A. K. Travel Time Estimation and Short-Term Prediction on Urban Arterial Networks Using Conditional Independence Graphs and State–Space Neural Networks. MS thesis. Michigan State University, 2006.

9. Sisiopiku, V. P., and N. M. Rouphail. Toward the Use of Detector Output for Arterial Link Travel Time Estimation: A Literature Review. In *Transportation Research Record 1457*, TRB, National Research Council, Washington, D.C., 1994, pp. 158–165.
10. Stathopoulos, A., and M. G. Karlaftis. A Multivariate State Space Approach for Urban Traffic Flow Modeling and Prediction. *Transportation Research Part C*, Vol. 11, 2003, pp. 121–135.
11. Lin, W. H., A. Kulkarni, and P. Mirchandani. Short-Term Arterial Travel Time Prediction for Advanced Traveler Information Systems. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, Vol. 8, No. 3, 2004, pp. 143–154.
12. Liu, H., H. J. van Zuylen, H. van Lint, and M. Salomons. Predicting Urban Arterial Travel Time with State–Space Neural Networks and Kalman Filters. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1968*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 99–108.
13. Saadaoui. Y. Prediction of Travel Time on Arterials with Random Events. MS thesis. School of Management, University of Tunis, 2010.
14. Zheng, F. Modeling Urban Travel Times. Ph.D. dissertation. Delft University of Technology, 2011.
15. Zheng, F., and H. van Zuylen. Estimating the Delay Distribution for Urban Trips Based on Sample Measurements. Presented at 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011.
16. van Zuylen, H., and F. Zheng. Modeling Variability of Urban Travel Times by Analyzing Delay Distribution for Multiple Signalized Intersections. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2259*, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 80–95.
17. Whittaker, J. *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester, 1990.
18. Haykin, S. *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, Upper Saddle River, N.J, 1999.

Artificial Intelligence and Transportation
Systems Modeling and Simulation

Transportation systems modeling and simulation is a very broad topic. In next two articles, the focus is more on using artificial intelligence (AI) to support modeling the supply side of the transportation system (a subsequent article will discuss AI applications for modeling the transportation system demand side). For the supply side, two articles are included. The first article, by Montasir Abbas, provides a brief overview of the agent-based approach to transportation systems modeling; the second article, by Shan Huang and Adel W. Sadek, surveys AI applications to the calibration of traffic simulation models.

Agent-Based Modeling and Simulation

MONTASIR ABBAS
Virginia Tech

Agent-based modeling and simulation (ABMS) is a relatively new approach to modeling complex systems by dividing them into a set of interacting autonomous agents. Agents have behaviors, often described by simple rules, and rules to change their behavior. By modeling agents individually, the full effects of the diversity that exists among agents in their attributes and behaviors can be observed as it gives rise to the behavior of the system as a whole. By modeling systems from the ground up (agent by agent), self organization and system evolution can often be observed in such models. ABMS techniques provide a rich framework for patterns, structures, and behaviors to emerge through the agent interactions. ABMS is arguably the only modeling technique that can simultaneously take into account the attributes and constraints imposed by geography of roadways, the impacts of continually evolving social networks, and the changes from individual decision making and learning in transportation modeling.

A preliminary review of literature on ABMS use in transportation reveals that existing efforts can be classified into two main threads:

1. Using ABMS to model ill-structured decision-making processes that cannot otherwise be accurately modeled. For example, modeling the evolution of driver's route-choice behavior, or safety-critical and aggressive driving. This is an area where rule-based or artificial intelligence (AI) ABMS techniques are suitable.

2. Using ABMS to better optimize the performance of the transportation system when solving the problem otherwise could be NP-complete or NP-hard. For example, optimizing a distributed control system. This is an area where cooperative agent techniques, such as game theory, are suitable.

These two main application areas are briefly surveyed below.

ABMS TO MODEL DECISION-MAKING PROCESSES

The literature contains an interesting use of ABMS techniques to model tactical driving and route choice behaviors.

Tactical Driving

Sukthankar et al. proposed a framework of autonomous agents that deal with specific aspects of tactical driving, where each agent makes independent driving decisions based on a subset of intelligent vehicle sensors (*1*). Population-based incremental learning (PBIL), a combination of genetic algorithm and competitive learning algorithms, managed the interaction of agents. Dresner and Stone presented an alternative mechanism for coordinating the movement of

autonomous vehicles through intersections (*2*). In this multiagent system, intersections use a new reservation-based approach built around a detailed communication protocol to accommodate human-driven vehicles in addition to autonomous vehicles. Abbas et al. developed artificial neural networks (ANNs) -based agents for modeling driver actions under perceived traffic state in car-following episodes (*3*). The neural agent is trained to be a clone of real driver through learning actions of an individual real driver behavior from naturalistic driving data. In a follow-up work, the authors trained the ANN agents using car-following episodes from one driver's naturalistic driving database to establish action rules that supersedes the Gazis–Herman–Rothery (GHR) car-following model (*4*).

Other research looked into the study of driver response and behavior for use with advanced traveler information system (ATIS). The belief–desire–intention (BDI) framework was developed with roots in philosophy and cognitive science, and is well described in the work of Bratman (*5*) and Bonsall and Parry (*6*) on rational agents. The BDI framework models a rational agent with bounded resources, limited understanding, and incomplete knowledge of what happens in its environment. These agents are most suitable to model the learning process of a human when selecting plans and evaluating their value to gain mode experience (*7, 8*). The BDI agents can be designed to vary their balance between reactive and deliberative behavior by changing the amount of time allowed for deciding what to do next. The BDI approach has been shown to be well suited to modeling different types of behaviors and has been successfully adopted in a number of fields such as tactical decision making in military operations and air traffic management (*9*).

An interesting work on human psychology and its link with tactical tasks was done by Maag et al. (*9*). In their work on a driver simulator study, a cognitive–emotional model for driver behavior is based on four main components: detection of interactions between drivers, calculation of emotional response, influencing driving style, and changing of driver's maneuvers. The goal of the research was to create anger in the driver and to see how this affects the behavior of the driver, particularly how this may cause aggressive driving behaviors (*10*). Hill and Boyle looked into the stress levels experienced by drivers during different driving conditions (*10*). The data for this research was gathered from a national survey that asked for a rating of stress (from one to seven) during different driving scenarios. The results showed that the distribution of stress levels was skewed higher for weather-related conditions and interactions with other drivers. Rothengatter discussed the differences between drivers and the need for more research into the psychology of drivers (*11*). Mesken researched the occurrence and magnitude of different emotions while driving. The results showed that the most common emotions were happiness, anger, and anxiety as reported by the participants (*12*). Matthews researched and found that the BDI is correlated to the effect or experience of stress in individual drivers or stress vulnerability (*13*).

Route Choice

Other work expanded on the cognitive process of individual drivers using the BDI framework to evaluate the impacts of the information from exogenous sources on the driver intentional reasoning system. Rossetti et al. applied a BDI framework to assess drivers' decision making in an extended version of the microscopic model DRACULA (*14, 15*). The BDI framework allowed driver agents to have autonomy in their travel decisions based on knowledge acquired from previous experiences. Wahle et al. used ABMS to analyze the impact of real-time

information in a two-route scenario (*16*). The driver model in their work was based on driving and route choice behavior and navigation layers. The results showed that information influenced route choice behavior of dynamic agents significantly in terms of concentration of drivers on the recommended routes which is intrinsic to many systems and leads to a negative impact on traffic pattern.

Dia proposed an agent-based approach to model individual driver behavior under the influence of real-time traffic information, where the traveler responds to real time information about traffic conditions, accident delays, road work, and route guidance (*17*). Agent behavior parameters including individual driver characteristics, knowledge, and preferences are identified through the responses of a survey implemented from a congested corridor in Brisbane, Australia. Xi et al. introduced a route guidance system based on an agent network (*18*). Dynamic vehicle route guidance system is made up of four main parts: information collection system, agent network, communication system, and information reception devices. A simulation study was based on PARAMICS and a route selection table for each driver agent was built and updated.

DISTRIBUTED MAS AND TRANSPORTATION
SYSTEM PERFORMANCE OPTIMIZATION

The literature shows situations where real-time simulation and off-line optimization were combined in a MAS framework, such as the work by Park and Kim (*19*). Fei-Yue and Cheng-Hong used agent-based control for development, management, and operation of intelligent network-enabled devices by transforming control algorithms to control agents (*20*). Fei-Yue broke down the control algorithms into task-oriented distributed control agents (*21*). The operation of the overall network is based on situation assessment, arbitration, and control fusion of decisions of more than one executing agent. In more recent work, Fei-Yue described an artificial, computational, parallel (ACP) approach to solve complex transportation systems modeling using parallel execution for control management with social and behavioral dimensions for operation (*22*).

Chong et al. presented an agent-based reinforcement learning framework to train controller agents to take appropriate actions according to traffic states. The neuro–fuzzy actor–critic reinforcement learning method was applied at isolated intersections. The control agents were trained to get knowledge of traffic states after the learning process and determine the optimal phase durations to minimize vehicle delay at a given intersection (*23*). A paper by Kosonen presents the signal control system based on real-time simulation, multiagent control scheme, and fuzzy inference. In the study, each signal negotiates with other agents (signals) to make decisions about the control strategies (*24*). In the approach presented by Ossowski, when a new acquaintance enters the system, agents are informed about its capabilities and, if the newcomer executes some signal plans that are previously unknown, they are added to the agents' knowledge bases (*25*).

The paper by Min Chee et al. presents an approach in which the large-scale traffic signal control problem is divided into various sub-problems and each sub-problem is handled by an agent who has a fuzzy neural decision-making ability (*26*). The higher-level agents undo any wrong decisions that are made by the lower-level agents. In order for the multiagent architecture to adapt itself continuously to the dynamically changing problem domain, a multistage online learning process for each agent is implemented involving reinforcement learning, learning rate,

and weight adjustment as well as dynamic update of fuzzy relations using an evolutionary algorithm (*26*). A paper by Roozemond investigates the applicability of the autonomous intelligent agents in urban traffic control (UTC) (*27*).The proposed system adapts itself to the changing environment automatically based on internal rules. This model is based on intelligent traffic signaling agents (ITSA) and authority agents which makes it possible to respond to real time traffic changes (*27*).

Traffic Control Integration with Route Selection

Another important application for ABMS is the integration of urban traffic signal control (UTSC) system and dynamic route guidance (DRG) system. Ruimin and Qixin used ABMS system in their paper as an example to discuss the architecture and interactive mechanism of agent models (*28*).

Weyns et al. used delegate multiagent systems for anticipatory vehicle routing to avoid traffic congestion (*29*). Individual vehicles are represented by agents to explore alternative routes in the environment on behalf of the vehicles. Cooperation between agents has been addressed using multiple methods. Adler et al. proposed a solution method based on the negotiation principle between agents that represent network managers, information service providers, and drivers equipped with route guidance systems (*30*). The proposed framework established the cooperative behavior of multiagents in the natural extension of the national ITS architecture. Adler et al. modeled the supplyside, real-time control over the transportation network through an agent-based distributed hierarchy of system operators (*31*). Allocation of network capacity and distribution of traffic advisories are performed by agents that act on behalf of information service providers. Drivers that need preferences are represented by agents embedded in intelligent in-vehicle route guidance systems. Balbo and Pinson implemented a multiagent system for public transportation system management (*32*). A multiagent decision support system (MADSS) was used by human regulators in order to manage bus lines by detecting bus delays and advances from the inconsistency of positioning data sent by buses to a central operator. Belmonte et al. also applied a multiagent approach building decision support system for bus fleet management problems (*33*). The system has been implemented on an agent-oriented tool, implemented in Java and FIPA-compliant (Jade).

Game Theory in ABMS

Mandiau et al. proposed behavior-based decision matrices for coordination between agents in an urban traffic simulation environment (*34*). A multiagent coordination mechanism was applied to intersection simulation situations. Driver agents are simulated, based on the principles of game theory, at intersections to avoid gridlocks and safety critical events. Decision matrices were built with an objective of maximizing gains and keeping interactional behavior to avoid possible unwanted situations. The theoretical results showed that agent-based simulation was best suited for a reasonable number of agents in conflict situations. Wangermann et al. described the role of the principle of negotiation in coordinating multiagent systems (*35*). He modeled intelligent agents as rule-based expert systems whose side effects are the procedural and reflexive functions of the agent. He also used multiattribute utility theory and decision trees for analyzing the behavior of agents (*35*).

Logi et al. provided a multiagent architecture for the provision of real-time decision support to traffic operations center personnel to help in coordinated, interjurisdictional traffic congestion management (*36*). Hernández et al. compared two multiagent systems that perform decision support for real-time traffic management in the urban motorway network (*37*). Van et al. presented a test bed for multiagent control systems in road traffic management in compliance with FIPA standards which allows transfer of the traffic-managing multiagent system to a real-world application. For the test bed, he used rule-based and Bayesian intelligence models to allow easy modeling of the business logic of the traffic engineers (*38*).

FUTURE OF ABMS IN TRANSPORTATION

The application of ABMS to transportation problems is an area that is still in its infancy. In the future, one may expect to see many more applications of ABMS in transportation. This is especially true given that ABMS is the ideal tool to use to model complex systems, and given that the transportation system, where thousands of different types of agents interact on a continuous basis, is a very good example of a complex system. One of the most challenging technical aspects of ABMS is the identification of worthy aspects or problems for application of the technique—and not only applying the tools "because we can." With the possibility of applying ABMS in existing optimization and simulation tools, translation of data and information or structure from one model to another will also be a significant challenge.

REFERENCES

1. Sukthankar, R., S. Baluja, and J. Hancock. Multiple Adaptive Agents for Tactical Driving. *Applied Intelligence*, Vol. 9 (compendex), 1998, pp. 7–23.
2. Dresner, K., and P. Stone. A Multiagent Approach to Autonomous Intersection Management. *Journal of Artificial Intelligence Research*, Vol. 31 (compendex), 2008, pp. 591–656.
3. Abbas, M., L. Chong, and A. Medina. Neural-Agent Driver Behavior Modeling Using Naturalistic Data. *Proc., 13th International IEEE Conference on Intelligent Transportation Systems*, Institute of Electrical and Electronics Engineers, Inc., 2010.
4. Chong, L., M. M. Abbas, and A. Medina. Simulation of Driver Behavior with Agent-Based Back-Propagation Neural Network. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2249*, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 233–241.
5. Bratman, M. E. Intentions, Plans and Practical Reasoning. Harvard University Press, Cambridge, Mass., 1987.
6. Bonsall, P., and T. Parry. Using an Interactive Route-Choice Simulator to Investigate Drivers' Compliance with Route Guidance Advice. In *Transportation Research Record 1306*, TRB, National Research Council, Washington, D.C., 1991, pp. 59–68.
7. Dia, H. An Agent-Based Approach to Modelling Driver Route Choice Behaviour Under the Influence of Real-Time Information. *Transportation Research Part C*, Vol. 10, No. 5–6, 2002, pp. 331–349.
8. Busetta, P., R. Ronnquist, A. Hodgson, and A. Lucas. Intelligent Agents—Components for Intelligent Agents in Java. 1999. Available at http://www.agent-software.com. Accessed December 28, 2000.
9. Maag, C., C. Mark, and H.-P. Krüger. Development of a Cognitive-Emotional Model for Driver Behavior, in Agent and Multi-Agent Systems: Technologies and Applications (P. Jedrzejowicz et al., eds.), Springer Berlin/Heidelberg, 2010, pp. 242–251.

10. Hill, J. D., and L. N. Boyle. Driver Stress as Influenced by Driving Maneuvers and Roadway Conditions. *Transportation Research Part F*, Vol. 10 (compendex), 2007, pp. 177–186.
11. Rothengatter, T. Psychological Aspects of Road User Behaviour. *Applied Psychology: An International Review-Psychologie Appliquee-Revue Internationale*, Vol. 46, No. 3, 1997, pp. 223–234.
12. Mesken, J., M. P. Hagenzieker, T. Rothengatter, and D. de Waard. Frequency, Determinants, and Consequences of Different Drivers' Emotions: An On-the-Road Study Using Self-Reports, (Observed) Behaviour, and Physiology. *Transportation Research Part F*, Vol. 10, No. 6, 2007, pp. 458–475.
13. Matthews, G., A. Guxin, and Y. Ozeki. Individual Differences in Driver Stress Vulnerability in a Japanese Sample. *Ergonomics*, Vol. 42 (compendex), 1999, pp. 401–415.
14. Rossetti, R. J. F., S. Bampi, R. Liu, D. Van Vliet, and H. B. B. Cybis. An Agent-Based Framework for the Assessment of Drivers' Decision-Making in Intelligent Transportation Systems. *Proc., IEEE*, 2000.
15. Rossetti, R. J. F., R. H. Bordini, A. L. C. Bazzan, S. Bampi, R. Liu, and D. Van Vliet. Using BDI Agents to Improve Driver Modelling in a Commuter Scenario. *Transportation Research Part C*, Vol. 10, No. 5–6, 2002, pp. 373–398.
16. Wahle, J., A. C. Bazzan, F. Klügl, and M. Schreckenberg. The Impact of Real-Time Information in a Two-Route Scenario Using Agent-Based Simulation. *Transportation Research Part C*, Vol. 10, No. 5–6, 2002, pp. 399–417.
17. Dia, H. An Agent-Based Approach to Modelling Driver Route Choice Behaviour Under the Influence of Real-Time Information. *Transportation Research Part C*, Vol. 10, No. 5–6, 2002, pp. 331–349.
18. Shi, X., J. Xu, Y. Xu, and J. Song. A Simulation Study on Agent-Network Based Route Guidance System. *Proc., 2005 IEEE in Intelligent Transportation Systems*, 2005.
19. Park, K., and W. Kim. A Systolic Parallel Simulation System for Dynamic Traffic Assignment: SPSS-DTA. *Expert Systems with Applications*, Vol. 21, No. 4, 2001, pp. 217–227.
20. Fei-Yue, W., and W. Cheng-Hong. Agent-Based Control Systems for Operation and Management of Intelligent Network-Enabled Devices in Systems, Man, and Cybernetics. Presented at IEEE International Conference on Intelligent Transportation Systems, 2003.
21. Fei-Yue, W. Agent-Based Control for Networked Traffic Management Systems. *Intelligent Systems*, IEEE, Vol. 20, No. 5, 2005, pp. 92–96.
22. Fei-Yue, W. Toward a Revolution in Transportation Operations: AI for Complex Systems. *Intelligent Systems*, IEEE, Vol. 23, No. 6, 2008, pp. 8–13.
23. Chong, L., and M. Abbas. Neuro-Fuzzy Actor Critic Reinforcement Learning for Determination of Optimal Timing Plans. *Proc., 13th International IEEE Conference on Intelligent Transportation Systems*, 2010.
24. Kosonen, I. Multi-Agent Fuzzy Signal Control Based on Real-Time Simulation. *Transportation Research Part C*, Vol. 11, No. 5, 2003, pp. 389–403.
25. Ossowski, S., J. Cuena, and A. García-Serrano. A Case of Multiagent Decision Support: Using Autonomous Agents for Urban Traffic Control. *Progress in Artificial Intelligence—IBERAMIA* (H. Coelho, ed.), Springer Berlin/Heidelberg, 1998, pp. 468–469.
26. Min Chee, C., D. Srinivasan, and R. L. Cheu. Cooperative, Hybrid Agent Architecture for Real-Time Traffic Signal Control. *Systems, Man and Cybernetics, Part A: Systems and Humans*, Vol. 33, No. 5, 2003, pp. 597–607.
27. Roozemond, D. A. Using Intelligent Agents for Pro-Active, Real-Time Urban Intersection Control. *European Journal of Operational Research*, Vol. 131, No. 2, 2001, pp. 293–301.
28. Ruimin, L., and S. Qixin. Study on Integration of Urban Traffic Control and Route Guidance Based on Multi-Agent Technology. *Proceedings of IEEE Intelligent Transportation Systems*, 2003.
29. Weyns, D., T. Holvoet, and A. Helleboogh. Anticipatory Vehicle Routing Using Delegate Multi-Agent Systems. *Intelligent Transportation Systems Conference*, IEEE, 2007.

30. Adler, J. L., and V. J. Blue. A Cooperative Multi-Agent Transportation Management and Route Guidance System. *Transportation Research Part C*, Vol. 10, No. 5–6, 2002, pp. 433–454.

31. Adler, J. L., G. Satapathy, V. Manikonda, B. Bowles, and V. J. Blue. A Multi-Agent Approach to Cooperative Traffic Management and Route Guidance. *Transportation Research Part B*, Vol. 39 (compendex), 2005, pp. 297–318.

32. Balbo, F., and S. Pinson. Toward a Multi-Agent Modelling Approach for Urban Public Transportation Systems. *Engineering Societies in the Agents World II* (A. Omicini, P. Petta, and R. Tolksdorf, ed.), Springer Berlin/Heidelberg, 2001, pp. 160–174.

33. Belmonte, M. V., J. L. Pérez-de-la-Cruz, and F. Triguero. Ontologies and Agents for a Bus Fleet Management System. *Expert Systems with Applications*, Vol. 34, No. 2, 2008, pp. 1351–1365.

34. Mandiau, R., A. Champion, J.-M. Auberlet, S. Espié, and C. Kolski. Behaviour Based on Decision Matrices for a Coordination Between Agents in a Urban Traffic Simulation. *Applied Intelligence*, Vol. 28 (compendex), 2008, pp. 121–138.

35. Wangermann, J. P., and R. F. Stengel. Principled Negotiation Between Intelligent Agents: A Model for Air Traffic Management. *Artificial Intelligence in Engineering*, Vol. 12, No. 3, 1998, pp. 177–187.

36. Logi, F., and S. G. Ritchie. A Multi-Agent Architecture for Cooperative Inter-Jurisdictional Traffic Congestion Management. *Transportation Research Part C*, Vol. 10, No. 5–6, 2002, pp. 507–527.

37. Hernández, J. Z., S. Ossowski, and A. García-Serrano. Multiagent Architectures for Intelligent Traffic Management Systems. *Transportation Research Part C*, Vol. 10, No. 5–6, 2002, pp. 473–506.

38. Van Katwijk, R. T., P. van Koningsbruggen, B. De Schutter, and J. Hellendoorn. Test Bed for Multiagent Control Systems in Road Traffic Management. Delft University of Technology, 2005.

Artificial Intelligence and Microscopic Traffic Simulation Models
Applications to Parameter Calibration and Origin–Destination Estimation

SHAN HUANG
ADEL W. SADEK
University at Buffalo, the State University of New York

In recent years, the use of microscopic traffic simulation (microsimulation hereafter) has become increasingly popular in transportation modeling. After creating a crude microsimulation model, an indispensible task is to tune up the model so it can accurately represent the real data. This task can be divided into two sub-tasks, namely parameter calibration, i.e., to adjust the parameter values of the model, and origin–destination estimation (ODE), i.e., to adjust the origin–destination (O-D) matrices representing the travel demand driving the simulation model.

The tune-up process of microsimulation models is often very time consuming and difficult. Moreover, the simulation parameters cannot be directly observed from the field data. This thus calls for a systematic process to help adjust the values of those parameters to allow the microsimulation model to better represent reality. On the other hand, O-D matrices are typically compiled from travel survey data. Given that it is practically impossible to survey every traveler, sampling techniques are always used to reduce the cost, which introduce another source of uncertainty. As a result, the O-D matrices always needed to be adjusted based on field counts, even within the context of a four-step planning model and a static traffic assignment. For microsimulation, the problem is more complex, because in that case, the dynamic demand or the time-dependent O-D matrices need to be estimated or adjusted.

As can be seen from the above, both calibrating the parameters of a microsimulation model or estimating the travel demand for deriving the model can be formulated as a search or optimization problem. Microsimulation models, however, are stochastic and discontinuous, and hence too complex to be mathematically formulated. At the same time, simulation-based optimization problems formulated to either calibrate the parameters or estimate the demand require running the simulation model for each evaluation of the objective function. The high computational cost associated with micro-simulation models is another obstacle to researchers, making it impractical to use an exhaustive search algorithm to try out all the possible combinations. To solve these two difficult problems, artificial intelligence (AI) methods have been researched and applied.

PARAMETER CALIBRATION

The parametric calibration problem is less complex than the ODE problem because its search space is much smaller. Genetic algorithms (GAs) offer several advantages in solving this problem. GAs do not require gradient information, are rather robust, and can overcome the combinatorial explosion of the simulation model calibration problem. Cheu et al. spearheaded a GA application for finding a suitable combination of CORSIM (FRESIM) microsimulator

parameter values based on data collected over a 5.8-km segment of an expressway in Singapore (*1, 2*). The parameters to be calibrated were coded into chromosomes in the form of a series of decimal numbers, and a simple GA was used to search for a best combination that can minimize the error between simulated traffic volumes and field volumes. Similarly, Lee at al. presented another application of GA for calibrating two Paramics microsimulator (*3*) parameter values based on data collected over a 1-mi segment of a freeway in California (*4*). However, both applications showed limited reduction of discrepancies between simulated and observed counts.

Ma and Abdulhai implemented GAs on a larger Paramics model in downtown Toronto, Canada, which had 1,270 links and 470 nodes (*5*). Their research involved the search for five Paramics parameter values. Other than the simple GA, three other GAs were also tested in that research, namely steady-state GA, crowding GA, and incremental GA. Their results showed that link flow error was reduced to 46% from the noncalibrated 51%.

Schultz and Rilett applied GA to calibrate 11 parameter values for a CORSIM model of a 13.9-mi (22.4-km) section of Interstate 10 in Houston, Texas, that had 14 on-ramps and 13 off-ramps, with five automatic vehicle identification (AVI) stations that divide the facility into four links (*6*). The parameters to be calibrated included 10 driver behavior parameters, specifically car-following sensitivity factors corresponding to 10 driver types, along with a car-following constant used in CORSIM's PITT car-following model. A unique feature of that study is that the researchers assumed that the 10 driver behavior parameters were following a lognormal distribution or a truncated normal distribution, and thus avoided the need to calibrate the 10 different parameters, but instead only needed to calibrate the two parameters of the probability distribution itself. This approach greatly reduced the search space. The simulation results were calibrated using both volume and travel time datasets that were observed in the field. The results showed a roughly 20% improvement in the objective function, and indicated that both the lognormal and normal distributions were effective at modeling observed conditions.

Kim and Rilett applied GA to calibrate two microsimulation models (CORSIM and TRANSIMS) separately (*7, 8*). For each model, two test beds were modeled. The first test bed was a 13.9-mi (22.4-km) section of highway I-10 in Houston (the same test bed referred to in the previous paragraph). The second test bed was a 23-km corridor along US-290 that had 12 on-ramps, 12 off-ramps, and seven AVI stations. Nineteen parameter values for CORSIM and three for TRANSIMS were calibrated. The results showed a good convergence of GA for both models.

Park and Qi presented another calibration method based on a VISSIM model containing one actuated signalized intersection (*9, 10*). In their research, eight critical parameters and their acceptable ranges were first identified by a Latin Hypercube Design algorithm and a one-way analysis of variance. Following that, GA was applied to find the optimal parameter values. The objective function was based on the discrepancy between simulated and measured travel times. The simulation results were validated based on three criteria: (*a*) comparison of distribution of 100 VISSIM simulation outputs, (*b*) paired *t*-test, and (*c*) animations. The results show that the calibrated parameters generated performance measures that were more representative of the field conditions than the default parameters.

Ma et al. developed a GA, a simultaneous perturbation stochastic approximation (SPSA) algorithm, and a trial-and-error iterative adjustment (IA) algorithm, and applied all of them to the calibration of four Paramics parameters in a model of SR-99 in Sacramento, California, that was a congested 24-mi freeway section (*11*). Instead of calibrating on traffic volume or travel time, this study used capacity and critical occupancy. Since they were not influenced by traffic volumes, an accurate O-D demand matrix was therefore not required. Their results indicated that

the SPSA algorithm can reach the same level of accuracy with considerably less CPU time than GA and IA, while GA managed to reach the best solution.

Ciuffo et al. developed an OptQuest–Multistart algorithm to calibrate eight parameters in an Aimsun model of a 11-km congested Italian freeway section (*12*, *13*). Field traffic counts and travel speeds were separately used as the calibration criterion, which both showed acceptable result qualities.

Smith et al. applied GA to calibrate five parameter values for a large-scale Paramics model of Chittenden County, Vermont, an area of about 1,400 km^2 (540 mi^2) (*14*). In their calibration, the simulated hourly volumes at the 405 intersection approaches were extracted from Paramics output and compared against the field data, and the average GEH value for each approach was calculated as the objective function. GA successfully reduced the average GEH from more than 6.5 to 5.91. Cunha et al. used GA to calibrate 40 parameter values for a CORSIM model and to calibrate 32 parameter values for an INTEGRATION model (*15*, *16*). A GPS receiver was installed in trucks to collect kinematic speed and position data along a section of approximately 10 km on a divided highway in Brazil. Data including total truck mass, nominal engine power, axle number, and configuration data, were collected at a mobile weigh station. From those data, performance curves were constructed for all vehicles in the sample. The mean absolute error ratio between simulated and observed performance curves was used as the objective function. The results showed the feasibility of using GA to calibrate both CORSIM and INTEGRATION models.

ORIGIN–DESTINATION ESTIMATION

For microsimulation models, ODE is a more complex problem than parameter calibration, due to its much larger search space. For parameter calibration, it can be clearly seen from the last section that the number of parameters to calibrate seldom exceeds 50. However, the number of unknowns (O-D pairs) in a microsimulation ODE problem ranges from to 100 (very small scale) to more than 1 million (large scale).

Kattan and Abdulhai developed a parallel GA framework to solve the ODE problem (*17*). A mesoscopic simulation model, namely Dynasmart-P (*18*), was used in that study. In their GA, the O-D matrices were first shaped into a vector form. Then the ODE problem was formulated as a generalized least-square estimation to minimize two objective functions: (*a*) the deviation between field data and the simulated values and (*b*) the deviation between the a priori demand matrix and the unknown time-dependent demand. A weighted formulation was used to combine the two sets of deviations. The GA was augmented with parallelization to improve the efficiency. This approach was evaluated on a small-scale real network in Toronto that consisted of 10 zones, 30 nodes, and 50 links. The estimation process included four estimation intervals, making the total O-D pairs to estimate equal to 400. The SPSA algorithm was compared with GA. The results showed that GA outperformed SPSA in terms of the objective function value.

Vaze also compared the performances of SPSA and a GA in solving the ODE problem using a MITSIMLab model (*19*, *20*). For each algorithm, a case that used multiple data sources (including AVI information) was compared with another case that used only loop detector data. Their methodology was first tested using a small synthetic study network, and then applied to a real traffic network in New York with 1,767 links, 825 nodes, and 482 O-D pairs. The total simulation period was divided into eight intervals. For each network, two experiments were

performed. In the first experiment, the parameter calibration and the ODE were performed sequentially. In the second experiment they were performed jointly. SPSA slightly outperformed GA for the small network, and was chosen for the large-scale case study. For the large-scale case, the total number of unknowns for the case of joint calibration was 3,856 O-D flows + 2,564 segment capacities + 50 speed/density parameters = 6,470 parameters. The results also showed that the additional AVI information had helped the joint calibration.

Huang et al. applied a parallel GA to solve large-scale ODE problems based on the TRANSIMS model (*21*). Two case studies were considered: (*a*) a synthetic small-scale network with nine zones, 82 nodes, and 141 links and (*b*) a realistic medium-scale network of a university campus in New York with 56 zones, 109 nodes, and 131 links. For the medium-scale case, 24 time intervals were involved, making the total number of unknowns to be 75,264 O-D flows. The objective function was represented by the root mean square error (RMSE) between simulated and field-collected traffic volumes. The results indicated that GAs appeared to have a significant impact on improving the quality of the solutions. Specifically, for the medium-scale network, GA resulted in a dramatic 76% reduction in the model's RMSE.

CONCLUSIONS

As can be seen from the above quick review of the literature, the use of GAs and AI-based search algorithms to calibrate the parameters of a microsimulation model or estimate the dynamic demand needed to run the model is an area that has received significant attention from researchers recently. The results generally illustrate the ability of GAs to bring the model's results closer to field observations. However, among the challenges of using GA in that context, and, particularly for ODE, is the very large solution space of the problem. Innovative methods are needed to help reduce the solution space or to guide the search technique in an efficient fashion. With the continued interest in the use and development of large-scale microsimulation models, this area is likely to attract more attention in the future.

REFERENCES

1. MCTRANS. CORSIM: Microscopic Traffic Simulation Model. 2011. Available at http://mctrans.ce.ufl.edu/featured/tsis/Version5/corsim.htm. Accessed March 10, 2011.
2. Cheu, R. L., X. Jin, K. C. Ng, Y. L. Ng, and D. Srinivasan. Calibration of FRESIM for Singapore Expressway Using Genetic Algorithm. *Journal of Transportation Engineering*, 1998, pp. 526–535.
3. QUADSTONE. Quadstone Paramics: Cutting Edge Traffic Simulation Solutions. 2010. Available at http://www.Paramics-online.com/. Accessed November 29, 2010.
4. Lee, D.-H., X. Yang, and P. Chandrasekar. Parameter Calibration for PARAMICS Using Genetic Algorithm. Presented at 80th Annual Meeting of the Transportation Research Board, Washington, D.C, 2001.
5. Ma, T., and B. Abdulhai. Genetic Algorithm-Based Optimization Approach and Generic Tool for Calibrating Traffic Microscopic Simulation Parameters. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1800*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 6–15.
6. Schultz, G. G., and L. R. Rilett. Analysis of Distribution and Calibration of Car-Following Sensitivity Parameters in Microscopic Traffic Simulation Models. In *Transportation Research Record: Journal*

of the Transportation Research Board, No. 1876, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 41–51.

7. Kim, K.-O., and L. R. Rilett. Genetic Algorithm-Based Approach to Traffic Microsimulation Calibration Using ITS Data. Presented at 83rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2004.

8. TRANSIMS. TRANSIMS Opensource. 2010. Available at http://transims-opensource.net/. Accessed November 29, 2010.

9. Park, B., and H. Qi. Development and Evaluation of a Procedure for the Calibration of Simulation Models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1934*, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 208–217.

10. PTV. VISSIM: Multi-Modal Traffic Flow Modeling. 2010. Available at http://www.ptvamerica.com/software/ptv-vision/vissim/. Accessed November 29, 2010.

11. Ma, J., H. Dong, and H. M. Zhang. Calibration of Microsimulation with Heuristic Optimization Methods. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1999*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 208–217.

12. Ciuffo, B. F., V. Punzo, and V. Torrieri. Comparison of simulation-based and model-based calibrations of traffic-flow microsimulation models. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2088*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 36–44.

13. TSS. Aimsun 6: The Integrated Transport Modelling Software. 2010. Available at http://www.aimsun.com/site/content/category/1/32/53/. Accessed November 29, 2010.

14. Smith, M. C., A. W. Sadek, and S. Huang. Large-Scale Microscopic Simulation: Toward an Increased Resolution of Transportation Models. *Journal of Transportation Engineering*, Vol. 134, No. 7, 2008, pp. 273–281.

15. Cunha, A. L., J. E. Bessa, and J. R. Setti. Genetic Algorithm for the Calibration of Vehicle Performance Models of Microscopic Traffic Simulators. *Progress in Artificial Intelligence*, Vol. 5816, 2009, pp. 3–14.

16. Rakha, H., and B. Crowther. Comparison and Calibration of FRESIM and INTEGRATION Steady-State Car-Following Behavior. *Transportation Research Part A*, Vol. 37, 2003, pp. 1–27.

17. Kattan, L., and B. Abdulhai. Noniterative Approach to Dynamic Traffic Origin–Destination Estimation Using Parallel Evolutionary Algorithms. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1964*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 201–210.

18. MCTRANS. DYNASMART-P version 1.3.0. 2011. Available at http://mctrans.ce.ufl.edu/featured/dynasmart/. Accessed March 10, 2011.

19. Vaze, V. S., C. Antoniou, Y. Wen, and M. E. Ben-Akiva, M. Calibration of Dynamic Traffic Assignment Models with Point-to-Point Traffic Surveillance. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2090*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 1–9.

20. MIT. MITSIMLab. 2011. Available at http://mit.edu/its/mitsimlab.html. Accessed March 10, 2011.

21. Huang, S., A. W. Sadek, I. Casas, and L. Guo. Calibrating Travel Demand in Large-Scale Micro-Simulation Models with Genetic Algorithms: A TRANSIMS Model Case Study. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.

Ramp Metering

GEORGE X. LU
University of Vermont

HONGCHAO LIU
Texas Tech University

R amp metering is a control strategy that regulates the frequency of vehicles entering the freeway at entrance ramps. Operated by either single or systemwide stop-and-go type traffic signals, the objective of ramp metering is to maintain optimum control of freeway and prevent operational breakdowns through limiting the rates of entering vehicles at entrance ramps or freeway-to-freeway connector ramps. For both pretimed and traffic-responsive local and systemwide control strategies, the underlying idea of a ramp metering algorithm is to balance demand and capacity of the freeway through well-defined objective functions and constraint sets (*1*). In this regard, linear programming has been widely used since the emergence of ramp metering concept in 1960s. Since then, transportation problems have become increasingly complex as their scope of analysis expanded rapidly beyond the traditional domain. These problems are characterized by

- A large number of factors involved;
- Parametric associations among the factors are unfathomable;
- A huge amount of incomplete data is encompassed; and
- Many objectives and constraints are so intertwined that the priorities among the stakeholders are blurry (*2*).

Kosko stated: "As the complexity of a system increases, our ability to make precise and yet significant statements about its behaviors diminishes, and significance and complexity become almost mutually exclusive characteristics" (*3*). Given that complex problems are difficult to solve using conventional methodologies, there has been a growing interest in employing artificial intelligence (AI) paradigms to address transportation issues to improve operation, safety, and efficiency of transportation systems.

ARTIFICIAL INTELLIGENCE AND RAMP METERING

Fuzzy Logic Based Algorithms

Fuzzy logic algorithms appear well suited to ramp metering because they can utilize inaccurate or imprecise information and allow smooth transition between metering rates. Inputs and outputs are descriptive (e.g., "no congestion," "light congestion," and "medium congestion") to allow for imprecise data. Fuzzy logic systems use rule-based logic to incorporate human expertise. This way, it can balance several performance objectives simultaneously and consider many types of

information, such as traffic conditions downstream, occupancy, flow rate, speed, and ramp queue. These capabilities allow fuzzy logic to anticipate the problem and take temperate and corrective action before the congestion occurs.

Sasaki and Akiyama developed a fuzzy traffic control system on an urban expressway (*4–6*). The authors showed that control of an urban expressway depends upon a skilled operator's judgment and decisions. The premier goal of the Sasaki and Akiyama's research was to investigate the effectiveness of fuzzy logic-based models in describing the operator's judgment process. A simple fuzzy reasoning model for on-ramp control and its performance were presented in their papers. The resulted strategies include restricting the number of booths and closing the gates, which proved to be successful from the test on the Osaka–Sakai route of the Hanshin Expressway. A main conclusion of their study is that fuzzy logic-based traffic controller system can effectively describe the judgment process and consequently take the place of human operators.

Chen et al. (*7*) presented a fuzzy controller for freeway ramp metering that uses rules of the form: IF "freeway condition" THEN "control action." The controller has been designed to consider varied levels of congestion, a downstream control area, changing occupancy levels, upstream flows, and a distributed detector array in its rule base. Through fuzzy implication, the inference of each rule is used to the degree to which the condition is true. Using a dynamic simulation model of conditions at the San Francisco–Oakland Bay Bridge, the action of the fuzzy controller is compared to the existing "crisp" control scheme, and an idealized controller. Tests under a variety of scenarios with different incident locations and capacity reductions show that the fuzzy controller is able to extract 40% to 100% of the possible savings in passenger-hours. In general, the fuzzy algorithm displays smooth and rapid response to incidents, and significantly reduces the minute-miles of congestion.

Taylor et al. designed a fuzzy logic ramp-metering algorithm to overcome the limitations of conventional ramp-metering strategies (*8*). The fuzzy controller also demonstrated improved robustness, prevented heavy congestion, intelligently balanced conflicting needs, and is user friendly. The objective was to maximize total distance traveled and minimize total travel time and vehicle delay while maintaining an acceptable ramp queue level. A multiple ramp study site from the Seattle I-5 corridor was modeled and tested using the freeway simulation software, FRESIM. For five of the six testing sets, encompassing a variety of traffic conditions, the fuzzy controller outperformed the three other controllers tested.

Vukanovic and Ernhofer presented the ACCEZZ ramp metering algorithm which is an adaptive control approach based on fuzzy logic (*9*). Results from its calibration and validation as well as its evaluation using a microscopic simulation show how different control strategies can improve traffic conditions especially during peak hours. To evaluate the potential benefits of ramp metering, the ramp metering software TRANSRAMP has been developed, implemented, and tested at a demonstration site in Munich, Germany. Since its implementation, TRANSRAMP controls traffic at two consecutive on ramps successfully with the ACCEZZ algorithm.

Jiang et al. designed a fuzzy self-adaptive proportional–integral–derivative (PID) controller and applied it to freeway ramp metering (*10*). A traffic flow model to describe the freeway flow process is built first. Based on the model and in conjunction with nonlinear feedback theory, a fuzzy-PID ramp controller is designed. The ramp metering rate is determined by the PID controller whose parameters are tuned by fuzzy logic according to the density tracking error and error variation. Gauss and triangle curves are used for the membership functions of the fuzzy variables. Finally, the control system is simulated in MATLAB software and the result

features good dynamic and steady-state performance. The model is advantageous in achieving a desired traffic density along the mainline of a freeway, and making vehicles travel more efficiently and safely.

Neural Network-Based Algorithms

Because traffic flow varies in time and space, it is essential to identify traffic characteristics and patterns at various combinations of time and space to define appropriate control strategies. Several previous studies used artificial neural networks (ANNs) to control freeway traffic operations by ramp metering in order to maintain desired level of service at various times and locations (*11–13*).

Zhang and Ritchie proposed a nonlinear approach for designing local traffic-responsive ramp controls using ANNs (*14*). The problem is formulated as a nonlinear feedback control problem, where the system model is the well-known hydrodynamic model developed by Lighthill and Whitham (*15*) and Richards (*16*), the model's flow–density relationship is nonlinear, and the feedback nonlinear controllers are composed of one or a number of feed-forward neural networks. These neural network controllers are of integral or proportional-plus-integral type, and can be tuned online to achieve prescribed performance. Initial simulation results show that such an approach is promising.

Zhang et al. also developed a coordinated traffic-responsive ramp control strategy based on feedback control and ANNs (*17*). The proposed feedback control law is nonlinear and realized by a series of neural networks. The parameters of the neural networks are obtained through a nonlinear optimization procedure. Traffic simulations show that the proposed nonlinear ramp control strategy compares favorably against the well-known linear quadratic control strategy in reducing total travel times, particularly at situations where drastic changes in traffic demand and road capacity occur.

Wei proposed an augmented-type network which includes several basic modules intelligently affiliated according to traffic characteristics on the freeway (*18*). Inputs to neural network models are traffic states in each time period on the freeway segments while outputs correspond to the desired metering rate at each entrance ramp. The simulation outcomes indicate very encouraging achievements when the proposed neural network model is employed to govern the freeway traffic operations.

Bai et al. proposed a new method which is called neuro-fuzzy adaptive dynamic programming with eligibility traces [NFADP(λ)] (*19*). With the introduction of neuro-fuzzy and eligibility traces, the performance of ADP is greatly enhanced. First of all, the expert experience is introduced to ADP, therefore the convergence of ADP is greatly reinforced. Second, with the learning strategy revised, the training of action network is accelerated. In order to achieve multiple ramps metering control, special performance index function is established in NFADP(λ). Extensive simulation on a hypothetical freeway are carried out with NFADP(λ), compared to ALINEA as a stand-alone strategy. Simulation results indicate that NFADP(λ) have good performances in both alleviating stochastic variations of the traffic demand and congestion situations.

Bogenberger et al. proposed a nonlinear approach for designing traffic-responsive and coordinated ramp control using a self-adapting fuzzy system (*20*). A NFIS (adaptive neuro-fuzzy inference system) was used to incorporate a hybrid learning procedure into the control system. The traffic-responsive metering rate was determined every minute by the neuro-fuzzy control

algorithm. Coordination between multiple on-ramps is ensured by the integration of a common input into all ramp controllers upstream of a bottleneck and a periodical update of the fuzzy control system every 15 min by a hybrid learning procedure. The objective of the online tuning process of the fuzzy parameters is to minimize the total time spent. Therefore Payne's traffic flow model and a deterministic queuing model are integrated into the control architecture. To assess the impacts of the neuro-fuzzy ramp metering algorithm a section of 25 km was simulated with the FREQ model and compared with two other control scenarios. The results of the simulation of the neuro-fuzzy algorithm are very promising.

Agent-Based Algorithms

Researchers and practitioners have realized that single-agent systems, multiagent systems, and distributed AI are widespread attractive because they consider a social side of computer systems, ranging from human–computer interaction over distributed problem solving to the simulation of social systems. Multiagent principles, as discussed by Abbas in a previous section of this circular, are ideal for building a wide range of computer applications, ranging from quite simple e-mail supervisors to complex and time-critical systems such as traffic control (*21*).

      From the viewpoint of a freeway administrative agent, it is imperative to design an appropriate control mechanism for ramp metering such that the traffic entering the freeway does not incur overflow or underflow. Overflow means that the freeway is over utilized and accidents or congestion may easily occur. Underflow means a low utilization rate of the freeway, which is not cost effective. Hou et al. applied the iterative learning control approach to address the traffic density control problem in a macroscopic level freeway environment with ramp metering (*22*). The traffic density control problem is first formulated into an output tracking and disturbance rejection problem. Through rigorous analysis, it is shown that the iterative learning control method can effectively deal with this class of control problem and greatly improve the traffic response. Next, the iterative learning control is combined with error feedback in a complementary modular manner to achieve the output tracking and system robustness. The effectiveness of the new approach is further verified through case studies with intensive simulations.

      By modeling the separate instruments as intelligent agents, it might be possible to tune the actions of the individual instruments through the agent concept of collaboration. Letting the individual instruments handle the most basic forms of coordination automatically might also relieve the traffic operator. Katwijk and Koningsbruggen demonstrated the aforementioned ideas using two simple examples: one in which consecutive ramp metering installations coordinate their actions to promote the flow at a downstream bottleneck and one in which traffic management instruments coordinate their actions to attain a common goal on the network level (*23*).

      Jacob and Abdulhai conducted research to develop a self-learning adaptive integrated freeway–arterial corridor control for both recurring and nonrecurring congestion (*24*). Reinforcement learning, an AI method for machine learning, is used to provide a single, multiple, or integrated optimal control agent for both recurrent and nonrecurrent congestion. The study involved testing the methodology in three different applications: freeway control using multiple ramp controls, integrated corridor control with a single ramp and variable message sign (VMS), and integrated corridor control with multiple ramps and VMS. The microsimulation tool Paramics, which has been used to train and evaluate the agent in an offline mode within a

simulated environment, is described. Results from various simulation case studies in Toronto are encouraging and have demonstrated the effectiveness and superiority of the technique.

CONCLUSIONS

This paper has surveyed the application of several AI paradigms to the problem of ramp metering. As discussed, the applied paradigms can be broadly classified into: (*a*) fuzzy logic-based algorithms; (*b*) neural network-based algorithms; and (*c*) agent-based algorithms. Results from the studies performed to evaluate such algorithms seem to indicate the promise of AI when applied in such a context.

REFERENCES

1. Papageorgiou, M., and A. Kotsialos. Freeway Ramp Metering: An Overview. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 3, No. 4, 2002, pp. 271–281.
2. Kikuchi, S. Artificial Intelligence in Transportation Analysis: Approaches, Methods, and Applications. *Transportation Research Part C*, Vol. 17, 2009, pp. 455–540.
3. Kosko, B. Fuzzy Logic. Art House, Helsinki, Finland, 1993, p. 363.
4. Sasaki, T., and T. Akiyama. Development of Fuzzy Traffic Control System on Urban Expressway. Proc., 5th IFAC/IFIP/IFORS International Conference in Transportation Systems, 1986, pp. 333–338.
5. Sasaki, T., and T. Akiyama. Fuzzy On-Ramp Control Model on Urban Expressway and Its Extension. *Transportation and Traffic Theory* (N. H. Gartner, and N. H. M. Wilson, eds.), Elsevier Science, New York, 1987, pp. 377–395.
6. Sasaki, T., and T. Akiyama. Traffic Control Process of Expressway by Fuzzy Logic. *Fuzzy Sets and Systems,* Vol. 26, 1988, pp. 165–178.
7. Chen, L., A. May, and D. Auslander. Freeway Ramp Control Using Fuzzy Set Theory for Inexact Reasoning. *Transportation Research Part A*, Vol. 24, No. 1, 1990, pp. 15–25.
8. Taylor, C., D. Meldrum, and L. Jacobson. Fuzzy Ramp Metering: Design Overview and Simulation Results. In *Transportation Research Record 1634*, TRB, National Research Council, Washington, D.C., 1998, pp. 10–18.
9. Vukanovic, S., and O. Ernhofer. Field Evaluation of the Fuzzy Logic Based Ramp Metering Algorithm ACCEZZ. Presented at 11th IFAC Symposium on Control in Transportation Systems, Delft, Netherlands, 2006.
10. Jiang, T., X. Liang, and X. Liang. Fuzzy Self-Adaptive PID Controller for Freeway Ramp Metering. Proc., International Conference on Measuring Technology and Mechatronics Automation, IEEE, 2009, pp. 570–573.
11. Stephanedes, Y. J., and X. Liu. Neural Networks in Freeway Control. Proc., Pacific Rim TransTech Conference, American Society of Civil Engineers, Vol. 1, 1993.
12. Papageorgiou, M. A Neural Network Approach to Freeway Network Traffic Control. *Transportation Systems: Theory and Application of Advanced Technology*, Vol. 2, Oxford, Pergamon Press, 1995.
13. Wei, C. H., and K. Y. Wu. Artificial Neural Network Model Applied to Freeway Ramp Metering Control. *Transportation Planning Journal,* Vol. 25, No. 3, 1996, pp. 335–356.
14. Zhang, H. M., and S. G. Ritchie. Freeway Ramp Metering Using Artificial Neural Networks. *Transportation Research Part C*, Vol. 5, No. 5, 1997, pp. 273–286.
15. Lighthill, M. J., and G. B. Whitham. On Kinematic Waves, II. A Theory of Traffic Flow on Long Crowded Roads. *Proc., Royal Society of London. Series A, Mathematical and Physical Sciences*, Vol. 229, 1955, pp. 317–345.

16. Richards, P. I. Models of Freeway Traffic and Control, Mathematical Models of Public Systems. *Simulation Councils Proceedings Series,* 1971, pp. 51–60.
17. Zhang, H. M., S. G. Ritchie, and R. Jayakrishnan. Coordinated Traffic Responsive Ramp Control via Nonlinear State Feedback. *Transportation Research Part C*, Vol. 9, No. 6, 2001, pp. 337–352.
18. Wei, C.-H. Analysis of Artificial Neural Network Models for Freeway Ramp Metering Control. *Artificial Intelligence in Engineering*, Vol. 15, 2001, pp.241–252.
19. Bai, X., D. Zhao, and J. Yi. Coordinated Multiple Ramps Metering Based on Neuro-Fuzzy Adaptive Dynamic Programming. Presented at International Joint Conference on Neural Networks, Atlanta, Ga., 2009, pp. 241–248.
20. Bogenberger, K., H. Keller, and S. Vukanovic. A Neuro-Fuzzy Algorithm for Coordinated Traffic Responsive Ramp Metering. *IEEE Intelligent Transportation Systems Conference Proceedings*, Oakland, Calif., 2001, pp. 94–99.
21. Schleiffer, R. Intelligent Agents in Traffic and Transportation. *Transportation Research Part C*, Vol. 10, 2002, pp. 325–329.
22. Hou, Z., J.-X. Xu, and J. Yan. An Iterative Learning Approach for Density Control of Freeway Traffic Flow via Ramp Metering. *Transportation Research Part C*, Vol.16, No.1, 2008, pp. 71–97.
23. Katwijk, R., and P. Koningsbruggen. Coordination of Traffic Management Instruments Using Agent Technology. *Transportation Research Part C*, Vol. 10, 2002, pp. 455–471.
24. Jacob, C., and B. Abdulhai. Automated Adaptive Traffic Corridor Control Using Reinforcement Learning: Approach and Case Studies. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1959*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 1–8.

Application Area 2:
Artificial Intelligence and Travel Demand Modeling

This section focuses on artificial intelligence (AI) applications for modeling the demand side of transportation systems and includes two contributions. The first, by Erel Avineri, presents the grounds that justify the application of AI techniques to travel behavior modeling and then describes the applications of a wide range of AI paradigms to travel behavior research, including fuzzy set theory, neural networks, genetic algorithms, and multiagent simulations. The second contribution, by David Reinke, traces some additional opportunities for AI in urban travel demand forecasting.

Travel Behavior Research

EREL AVINERI
University of the West of England, Bristol

WHAT MAKES ARTIFICIAL INTELLIGENCE RELEVANT AND APPROPRIATE FOR THE ANALYSIS AND MODELING OF TRAVEL BEHAVIOR?

Travel behavior has been an area of great interest to practitioners, researchers, and policy makers interested in the demand side of transport systems. The travel choices made by travelers have a direct impact on the performance of transport systems and networks. Moreover, our travel behavior generates both positive and negative effects on our wealth, health, and well-being. In addition, travel behavior arising from the choices of individuals is perhaps the most significant determinant of effectiveness of transport policies and schemes. It is therefore desirable to understand and predict how people make travel choices, how travel choices might affect the overall performance of the transport system, and how travelers' behavior can be influenced in order to make it more efficient, safe, and sustainable.

In the conventional modeling approach applied in transportation planning, the model has generally been subdivided into four stages: trip generation, trip distribution, mode choice, and traffic assignment. Travel choice models applied in these four stages are designed to emulate the behavior of travelers over time and space and to predict changes in system performance when influencing conditions are changed. Commonly, it is the behavior of individual travelers that is analyzed and modeled (although households, organizations, or other entities could also be modeled). Such models include the mathematical and logical abstractions of choice behaviors implemented in algorithms and computer software. Among the travel behaviors that are commonly modeled are route choices, mode choices (i.e., the decision to travel by car, public transport, cycling, etc.) and travel time choices (i.e., departure time), and their combinations. The behavioral assumptions applied in the analysis and modeling of travel choices can be traced back to economic theory. Discrete choice analysis is commonly applied in the analysis of travel choices and is largely based on the paradigms of random utility theory (*1–3*).

In addition and sometimes as an alternative to the traditional models of travel behavior, artificial intelligence (AI) methods and techniques have been applied to model and analyze the behavior of travelers. The use of soft computing methodologies is of particular interest to transport researchers and practitioners due to their ability to handle quantitative and qualitative measures and to efficiently solve problems that involve complexity, imprecision, and uncertainty—many times applying methods that have some similarities to cognitive mechanisms applied by individuals in the process of choice making and problem solving. Moreover, the characteristics and performances of transport systems, and many of the perceived attributes of the travel alternatives in an individual's choice set, cannot always be simply defined or described on the basis of crisp and quantitative evaluation of their main effects. Much of the traveler's decisions and behaviors take place under imprecision, uncertainty, and partial truth. Some objectives, criteria, and constraints involved in travel choices are often difficult to be measured by crisp values, and are thus often neglected by transport modelers. Moreover, issues of travel

behavior are complex and inexact, and the traditional models cannot deal effectively with travelers' ambiguities, uncertainties, and vagueness. Some of these modeling challenges can be addressed by AI methods, as illustrated in the next sections.

AI PARADIGMS, COGNITIVE PSYCHOLOGY, AND TRAVEL BEHAVIOR

As van Zuylen mentioned before in his introductory section to this circular, AI systems and research can be classified into two categories (*4*). "Strong" AI research is intended to produce machines with an intelligence that matches or exceeds that of human beings. AI, in its weak form, is less ambitious; it concerns itself more with the degree to which machines can demonstrate mechanisms that underlie human behavior, but do not necessarily have consciousness, personal identity, emotions, or mind. Both forms of AI sit alongside cognitive psychology (and other behavioral sciences) at the core of an interdisciplinary approach to understanding and modeling intelligent behavior.

The main modeling tools used for the analysis modeling and prediction of travel choices stem from neoclassical economics in which individuals are assumed to make choices that are rational, consistent, perfectly informed, and maximize their economic utility by trading off between costs and benefits (*5*). However, research in behavioral sciences, especially cognitive psychology, indicates that individuals' choices in a wide range of contexts deviate from the predictions of the simpler forms of economic theory. Some of these deviations are systematic, consistent, robust, and largely predictable. Evidence on systematic deviations from rational models has emerged from studies on financial behavior, consumer behavior, health behavior (e.g., Tversky and Kahneman (*6–8*), Ariely (*9*), and Thaler and Sunstein (*10*)], and more recently travel behavior [e.g., Avineri and Prashker (*11*, *12*), van de Kaa (*13*), and Avineri and Chorus (*14*)]. This is somewhat in conflict with the traditional model of travel choice, in which travel is seen as a "derived demand," rationalized by its economic context, and travelers are expected to act as rational human beings, and exhibit consistency and transitivity in their choices.

It is clear that both AI and cognitive psychology should not be seen as two separate approaches relevant to the study of travel behavior; combining one with the other could promote both approaches to their full potential. Much of the recent work in cognitive psychology and behavioral economics focuses on the cognitive biases and the cognitive bounded rationality of decision makers; a novel approach in the study of travel behavior would be to incorporate systematic deviations of travel choice from the rational models of utility maximization into models and simulations of travel behavior, rather than aiming to develop algorithms that are based on a normative model of decision making. To truly imitate the choices and behaviors of real humans it would be necessary to model the biases, flaws, and limitations in their processes of judging, inferencing, learning, and problem solving, as exhibited in their travel behavior. For example, the incorporation of a cognitive psychology notions related to human perception of uncertainty into a fuzzy-based model of travel choice is illustrated in Avineri (*15*). To fully address the more-ambitious strong AI paradigm, future research could explore the incorporation of affective factors that play a role travel choice decisions through a range of emotional states. However, current AI research and applications into the field of travel behavior is typically associated with the weak paradigm of AI.

FUZZY SETS THEORY

Because of uncertainty and variability of the supply and demand sides of the transport system, travel choice models mainly measure the uncertainty of the system, but not always attempt to capture the uncertainty in the mind of the traveler and its effect on travel choices. In order to make practical use of travel choice models in stochastic networks a link is required between objectively measurable uncertainty of the transport system and travelers' perception of that uncertainty. We can identify three key reasons why fuzzy sets theory might be relevant to applications in travel behavior. First, imprecision and vagueness are inherent to the traveler's cognitive model of behavior and choice. Second, in the transport environment, the information obtained by the traveler in the formulation of preferences, decision variables, constraints, and parameters is vague or not precisely measurable. Third, imprecision and vagueness as a result of perception errors, cognitive biases, and subjective opinion may further dampen the quality and quantity of available information. Hence, fuzzy sets can be used to bridge modeling gaps of normative and descriptive decision models in travel behavior research (and bring us a step closer to the strong paradigm of AI).

The fuzziness of perception reflects humans' limited cognitive abilities and finite ability of sensory organ to resolve detail, store, and process information. Fuzzy sets theory, as a paradigm to deal with difficulties that are related to the concepts that have vague boundaries, have been fruitfully applied to a range transportation problems (16). Because travel choice cannot be separated from human perception and decision processes, it makes a good domain for fuzzy theory applications.

Since the early 1990s, fuzzy sets were applied in the field of travel behaviors and much of these applications have been in the modeling of the route-choice decision-making process, which is an essential part of any traffic assignment model (17–21).
In the existing literature there are two main approaches to construct fuzzy utility functions. The first approach is based on fuzzy graphs and allows describing human preferences over the alternatives using fuzzy production (if–then) rules. The second approach implies construction of fuzzy set valued functions, in particular fuzzy number-valued functions. Accordingly, fuzzy models of travel choice can be roughly separated into two types (21); models of the first type mostly derive from the initial work of Lotan et al. (19, 20) and are based on fuzzy rules and on the classical tools of fuzzy control. They handle rules such as: "If travel time on route A is *very short*, and travel time on route B is *intermediate*, then the driver will *certainly* choose route A."

Such an inference mechanism is typically applied in situations where the premises or the consequences can be described in fuzzy rather than real terms (i.e., by fuzzy variables associated with membership functions). In such a fuzzy rule-based system, the confidence with which the data match the premise is calculated; if this premise confidence is at least equal to a specified threshold value, the rule is said to be fireable, i.e., activated. When a rule is fired, the consequent actions are carried out to extract the traveler decision. The confidence with which these actions are taken depends on both the premise confidence and the confidence placed in the rule itself; this net confidence is the fuzzy AND (or minimum) of the premise and the rule confidences, and is called the posterior confidence. The premise confidence and rule-firing threshold determine whether an instance of a rule is fireable; a rule's instance is fireable if the premise confidence equals or exceeds the threshold. The posterior confidence is the confidence with which the consequent is executed, and is normally the confidence value stored with any data made or modified by the rule. For practical reasons, many models require having the fuzzy outcome

reduced into a single and crisp value representation; a deffuzification process may be therefore applied on the fuzzy consequent (*22*).

The second type of a fuzzy model of travel choice is based on the evaluation of possibility theory or with comparison tools that can be demonstrated to be equivalent (*21*).

The first studies of fuzzy-based choice modeling illustrated the possibilities of fuzzy logic in solving problems using hypothetical (and sometimes rather simplistic) numerical examples. Observed behavior (or revealed preferences) obtained at laboratory experiments and field studies have been successfully applied to calibrate and validate fuzzy-based choice models.

Some studies suggested extending Wardrop's principle of user equilibrium (*23*) to accommodate principles of fuzzy sets in the formulization of the transport system's uncertainties, as perceived by the traveler; travelers' perception of generalized travel time can be modeled using a fuzzy number (*24*). Thus, based on the perceived generalized travel times of the different route alternatives, travelers are expected to choose a route which optimizes their fuzzy travel time. An alternative approach to solve the fuzzy user equilibrium or to apply a fuzzy-based approach in the assignment is described in Wang and Liao (*25*). Yet the relevance of such a modeling approach to the modeling of travel behavior is a matter of empirical evidence and validation studies.

Among their applications to the study of travel behavior, fuzzy sets were also applied in the modeling of mode-choice behavior (*26–28*), daily activity schedules (*29*) and parking behavior (*30*).

NEURAL NETWORKS

Neural networks have been widely applied to a wide range of transport problems that defy traditional modeling approaches (*31*). Some argue that the neural network approach was based on the assumption that there is a similarity between the process of traveler decision making and the problem-solving approach on neural computing. In principle, however, the neural network approach is less representative of real traveler decision making than a fuzzy rule-based model. The operation of a fuzzy rule-based model enables adjustment of rules to improve the overall performance of the model. The performance of a neural networks model can be adjusted only by re-training using alternative data (*32*). An adaptive neuro fuzzy inference system approach, where a fuzzy inference system can be represented with a neural network structure (*33*), may be considered as a methodology that combines both approaches of soft computing to travel behavior modeling. An early example of a hybrid model is described in Vythoulkas (*34*), where route-choice decision making is modeled by a fuzzy rule base, and a neural network approach is used for calibrating the parameters of the fuzzy model. The use of neural networks for trip generation, trip distribution, and modal split models was demonstrated in a range of works (*35–41*). Cantarella and de Luca developed a novel modeling approach and demonstrated how to apply neural networks in the analysis of mode choice (*42*). Their approach combines the use of multilayer feed-forward networks (MLFFNs) with elements of random utility models. The proposed approach was applied to two case studies. The MLFFNs model was calibrated against revealed preferences and its performance was compared with random utility models. The results showed that the MLFFNs may outperform random utility models when the values of mode shares are similar. However, the application of MLFFNs, although being effective in the prediction of travel choices, provides very little insight on travel behavior. Compared with the

traditional econometric approaches, its contribution to the understanding of the variables contributing to mode choices is rather limited, as it cannot provide clear interpretation of parameters (*42*).

Celikiglu observed that in most of the applications of neural networks to the modeling of travel behavior, the feed-forward back propagation neural network (FFBPNN) models or hybrid models of FFBPNNs were proposed (*43*). He argued that the FFBPNN algorithm has drawbacks that can lead the model to develop in an inaccurate direction, and proposed two alternative approaches for travel mode choice analysis: radial basis function neural network and generalized regression neural network.

The results reported in other studies highlight the potential of neural networks in the analysis of travel (mode) choices (*42, 43*). Following the recent development of alternative models of travel choice, such as fuzzy sets or neural network models, there is a growing need for a systematic approach to validate and compare choice models within a general protocol. De Luca and Cantarella argued in favor of such a systematic approach, as the commonly used indicators (such as rho-square statistics) are relevant only for utility-based models and provide only little insights into model effectiveness (*44*). They introduced benchmarking indicators to validate and compare choice models, among them the model efficiency (computational speed and memory requirements).

For additional discussion on the advantages and limitations of neural networks in modeling travel behavior, see the paper by Reinke featured in this circular.

GENETIC ALGORITHMS

Genetic algorithms have been applied to route-choice modeling (*45–47*). In order to examine the dynamic nature of a driver–network system through microsimulation tools, Nakayama et al. developed a theoretical model of drivers' cognition, learning, and route choice, assuming limitations in drivers' cognitive capabilities (*46*). In their work, a production system, which is a compilation of if–then rules, was formulated to represent alternative strategies in route choice. In the proposed system, drivers learn from experience and apply inductive reasoning. Such framework is adopted in the above study because it has its basis in cognitive psychology and is also computationally tractable. The decision rules applied in the driver's route choice process and their inferiority values are revised and updated by the application of genetic algorithms through reproduction, crossover, and mutation.

Another interesting application of genetic algorithms to travel behavior is illustrated in Pribyl and Goulias (*48*). The proposed approach is based on the belief that people and households with similar sociodemographic characteristics have similar travel patterns (*49, 50*). Applying a method based on k-medoid clustering, groups of households with similar activity patterns are identified and clustered. An improvement in searching for clusters was incorporated by using genetic algorithms (*48*).

MULTIAGENT SIMULATIONS

Microscopic models of travel behavior can be largely described as multiagent simulations. Agents, representing individual travelers, households, or other decision-making entities, are

maintained as artificial individual entities with individual attributes and individual states, and make individual decisions based on these attributes and states. Two agents, when submitted to the same situation, can therefore make fundamentally different decisions (*51*). For a review of agent-based approaches to model a range of transport problems, travel behavior among them, see Bernhardt (*52*).

Such a modeling framework therefore might be considered as relatively realistic in describing the heterogeneity and the complexity in large-scale real-life settings. However it can be generally argued that the common approach in the application of multiagent models in travel behavior is that it is not the agents that evaluate alternatives, make choices and learn, but the overall system. In that respect, most applications in this area fail to address the strong AI paradigm, and it can be argued that many multiagent simulations should not be considered as truly AI techniques. However there is a potential in using advanced computational techniques in multiagent simulations. In a state-of-art review of computational methods, Nagel and Marchal consider parallel computing as a mature technology that can be used to speed up individual modules of such simulations (although practical applications are somewhat rare) (*51*). Distributed AI, software agents, and peer-to-peer systems are among the emerging computational methods that have recently attracted attention by the research community.

The modeling and analysis of travel behavior are typically disaggregated, meaning that the models represent the choice behavior of an individual entity. However, individuals are influenced by significant others, people in their social networks, and people who have geographical and social proximity. The interactions within and between travelers and social groups, and some complex behaviors such as social norms, social imitation, and social learning, are not commonly formulated in travel behavior models. Although there is a growing interest in the study of social interactions in relation to travel behavior, there is not much experience with incorporating social interactions into travel behavior models. The framework of multiagent models, which focuses on studying the patterns of social interaction among a population of agents, has been applied in the study of travel behavior (*53*) and in the modeling of the diffusion of transport technologies (*54*). Sunitiyoso et al. used laboratory experiments to reveal the learning process of travelers when making repeated travel decisions; experimental settings, such as interaction between participants and flow of information, were controlled (*55*). This data was used as an input to a simulation experiment utilizing an agent-based model, simulating larger group sizes, longer time periods, and complex situational settings that cannot be explored in the laboratory environment. It can be generally argued that agent-based models and simulations that are based primarily on normative assumptions on travel behavior might not correspond closely to travelers' real-world contextual settings (*55*). Observed behaviors of "real" travelers could inform the modeling and simulation of artificial entities that represent such travelers. Such an approach can also help in ensuring the relevance validity of the results in representing a wider population, and can be used to investigate the potential effects of a transport measure prior to its implementation in practice.

CONCLUSIONS

Over the last two decades there has been a growing interest in applying AI to the study of travel behavior, leading to some successful applications and implementations. This paper reviewed a range of theories and techniques, such as fuzzy sets theory, artificial neural networks, genetic

algorithms, and agent-based models. Although it is by no mean a complete review of AI applications in the field of travel behavior, the author hopes that this paper has contributed to a better understanding of the potential use of AI methodologies in this field and provides the reader with a starting point when investigating the literature dealing with such applications.

REFERENCES

1. Domencich, T. A., and D. McFadden. *Urban Travel Demand: A Behavioral Analysis*. American Elsevier, New York, 1975.
2. Ben-Akiva, M. E., and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Mass., 1985.
3. Cascetta, E. Transportation Systems Engineering: Theory and Methods. *Applied Optimization Series*, No. 49, Kluwer Academic Publishers, Norwell, Mass., 2001.
4. Searle, J. Minds, Brains and Programs. *Behavioral and Brain Sciences*, Vol. 3, No. 3, 1980, pp. 417–457.
5. Avineri, E. On the Use and Potential of Behavioural Economics from the Perspective of Transport and Climate Change. *Journal of Transport Geography*, 2012.
6. Tversky, A., and D. Kahneman. Judgment Under Uncertainty: Heuristics and Biases. *Science*, Vol. 185, 1974, pp. 1124–1131.
7. Kahneman, D., and A. Tversky. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, Vol. 47, No. 2, 1979, pp. 263–291.
8. Tversky, A., and D. Kahneman. Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, Vol. 9, 1992, pp. 195–230.
9. Ariely, D. *Predictably Irrational: The Hidden Forces the Shape Our Decisions*. Harper-Collins, N.Y., 2008.
10. Thaler, R., and C. R. Sunstein. *Nudge: Improving Decisions about Health, Wealth and Happiness*. Yale University Press, New Haven, Conn., 2008.
11. Avineri, E., and J. N. Prashker. Violations of Expected Utility Theory in Route-Choice Stated Preferences: Certainty Effect and Inflation of Small Probabilities. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1894*, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 222–229.
12. Avineri, E., and J. N. Prashker. Sensitivity to Travel Time Variability: Travelers' Learning Perspective. *Transportation Research Part C*, Vol. 13, No. 2, 2005, pp. 157–183.
13. van de Kaa, E. J. Extended Prospect Theory. Findings on Choice Behaviour from Economics and the Behavioural Sciences and Their Relevance for Travel Behaviour. T2008/11. TRAIL Thesis Series, Netherlands, 2008.
14. Avineri, E., and C. Chorus. Recent Developments in Prospect Theory-Based Travel Behavior Research. *European Journal of Transport and Infrastructure Research*, Vol. 10, No. 4, 2010, pp. 293–298.
15. Avineri, E. Fuzzy Meaning of Reference-Based Perceptions in Travel Choice Modeling. Presented at 88th Annual Meeting of the Transportation Research Board, Washington D.C., 2009.
16. Kikuchi, S. Fuzzy Sets Theory Approach to Transportation Problems. In *Transportation Research Circular E-C113: Artificial Intelligence in Transportation: Information for Application*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 33–48.
17. Teodorović, D., and S. Kikuchi. Transportation Route Choice Model Using Fuzzy Inference Technique. *Proc., ISUMA '90: The First International Symposium on Uncertainty Modeling and Analysis*, IEEE Computer Society Press, College Park, Md., 1990, pp. 140–145.

18. Teodorović, D., and S. Kikuchi. Transportation Route Choice Model Using Fuzzy Inference Technique. *Proc., 1st International Symposium on Uncertainty Modeling and Analysis* (B. M. Ayyub, ed.), IEEE Computer Press, Los Alamitos, Calif., 1991, pp. 140–145.
19. Lotan, T., and H. N. Koutsopoulos. Models for Route Choice Behavior in the Presence of Information Using Concepts from Fuzzy Set Theory and Approximate Reasoning. *Transportation*, Vol. 20, 1993, pp. 129–155.
20. Lotan, T., H. N. Koutsopoulos, and Q. Yang. A Driving Simulator and Its Application for Modeling Route Choice in the Presence of Information. *Transportation Research Part C*, Vol. 2, No. 2, 1994, pp. 91–107.
21. Henn, V. Fuzzy Route Choice Model for Traffic Assignment. *Fuzzy Sets and Systems*, Vol. 116, No. 1, 2000, pp. 77–101.
22. Kosko, B. *Fuzzy Engineering*, Prentice Hall, 1997.
23. Wardrop, J. G. Some Theoretical Aspects of Road Traffic Research. *Proc., Institute of Civil Engineers, Part II,* Vol. 1, 1952, pp. 325–378.
24. Zhao, Z. The ε-Equilibrium in Transportation Networks. *Fuzzy Sets and Systems*, Vol. 68, 1994, pp. 195–202.
25. Wang, H.-F., and H.-L. Liao. User Equilibrium in Traffic Assignment Problem with Fuzzy N-A Incidence Matrix. *Fuzzy Sets and Systems*, Vol. 107, No. 3, 1999, pp. 245–253.
26. Teodorovic, D., and G. Pavkovic. The Fuzzy Set Theory Approach to the Vehicle Routing Problem When Demand at Nodes Is Uncertain. *Fuzzy Sets and Systems*, Vol. 82, 1996, pp. 307–317.
27. Cantarella, G. E., and V. Fedele. Fuzzy Utility Models for Analysing Mode Choice Behavior. Presented at European Transport Conference, 2003.
28. Sayed, T., and A. Razavi. Comparison of Neural and Conventional Approaches to Mode Choice Analysis. *Journal of Computing in Civil Engineering*, Vol. 14, No. 1, 2000, pp. 23–30.
29. Olaru, D., and B. Smith. Applying Neural Networks to the Interaction Between Land-Use and Transportation Networks: A Comparison with Multivariate Techniques. *Proc., 3rd International Conference: Decision Making in Urban and Civil Engineering,* London, 2002, pp. 343–350.
30. Dell'Orco, M., M. Ottomanelli, and D. Sassanelli. Modeling Uncertainty in Parking Choice Behavior. Presented at 82nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2003.
31. Ishak, S., and F. Trifiro. Neural Networks. *Transportation Research Circular E-C113: Artificial Intelligence in Transportation: Information for Application*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 17–32.
32. Hunt, J. G., and G. D. Lyons. Aspects of the Application of Artificial Neural Networks to Model Driver Decisions. *Proc., International Conference on Neural Network Applications in Transport*, VTT, Finland, 1995, pp. 15–34.
33. Jang, J.-S. R. ANFIS: Adaptive-Network-based Fuzzy Inference Systems. *IEEE Transaction on Systems, Man, and Cybernetics*, Vol. 23, 1993, pp. 665–685.
34. Vythoulkas, P. C. An Approach to Travel Behavior Based on the Concepts of Fuzzy Logic and Neural Networks. *Proc., Seminar G on Transportation Planning Methods: 22nd PTRC Transport Forum, Education and Research Services*, London, 1994, pp. 189–205.
35. Reggiani, A., and O. Tritapepe. Neural Networks and Logit Models Applied to Commuters' Mobility in the Metropolitan Area of Milan. *Neural Networks in Transport Systems* (V. Himanen, P. Nijkamp, and A. Reggiani, eds.), Ashgate, Brookfield, 1998.
36. Nijkamp, P., A. Reggiani, and T. Tritapepe. Modelling Inter-Urban Transport Flows in Italy: A Comparison Between Neural Network Analysis and Logit Analysis. *Transportation Research C*, Vol. 4, 1996, pp. 323–338.
37. Schintler, L. A., and O. Olurotimi. Neural Networks as Adaptive Logit Models. *Neural Networks in Transport Systems* (V. Himanen, P. Nijkamp, and A. Reggiani, eds.), Ashgate, Brookfield, 1998.
38. Shmueli, D., I. Salomon, and D. Shefer. Neural Network Analysis of Travel Behavior: Evaluating Tools for Prediction. *Transportation Research C*, Vol. 4, 1996, pp. 151–166.

39. Shmueli, D. Applications of Neural Networks in Transportation Planning. *Progress in Planning*, Vol. 50, 1998, pp. 141–204.
40. Kim, D. Neural Networks for Trip Generation Model. *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 4, 2001, No. 2, pp. 201–208.
41. Mozolin, M., J. C. Thill, and E. Lynn. Trip Distribution Forecasting with Multiplayer Perceptron Neural Networks: A Critical Evaluation. *Transportation Research B,* Vol. 34, 2000, pp. 53–73.
42. Cantarella, G. E., and S. de Luca. Multilayer Feed Forward Networks for Transportation Mode Choice Analysis: An Analysis and a Comparison with Random Utility Models. *Transportation Research C,* Vol. 13, No. 2, 2005, pp. 121–155.
43. Celikoglu, H. B. Application of Radial Basis Function and Generalized Regression Neural Networks in Non-Linear Utility Function Specification for Travel Mode Choice Modeling. *Mathematical and Computer Modelling*, Vol. 44, No. 7–8, 2006, pp. 640–658.
44. de Luca, S., and G. E. Cantarella. Validation and Comparison of Choice Models. *Travel Demand Management and Road User Pricing: Success, Failure and Feasibility* (W. Saleh and G. Sammer, eds.), Ashgate, U.K., 2009, pp. 37–58.
45. Sakai, A., S. Odagawa, and Y. Masumoto. Development of Route Calculation by Genetic Algorithm. Presented at Third Annual World Congress on Intelligent Transport Systems, Orlando, Fla., 1996.
46. Nakayama, S., and R. Kitamura. Route Choice Model with Inductive Learning. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1725*, TRB, National Research Council, Washington, D.C., 2000, pp. 63–70.
47. Nakayama, S., R. Kitamura, and S. Fujii. Drivers' Route Choice Rules and Network Behavior: Do Drivers Become Rational and Homogeneous Through Leaning? Presented at IATBR 2000 Conference, Sydney, Australia, 2000.
48. Pribyl, O., and K. G. Goulias. Simulation of Daily Activity Patterns Incorporating Interactions Within Households: Algorithm Overview and Performance. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1926*, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 135–141.
49. Stopher, P. R., and H. M. A. Metcalf. Household Activities, Life Cycle, and Role Allocation Tests on Data Sets from Boston and Salt Lake City. In *Transportation Research Record 1676*, TRB, National Research Council, Washington, D.C., 1999, pp. 95–102.
50. Bowman, J. L., and M. Ben-Akiva. Activity-Based Travel Forecasting. *Activity-Based Travel Forecasting Conference* (L. J. Engelke, ed.), Texas Transportation Institute, Austin, 1997, pp. 3–36.
51. Nagel, K., and F. Marchal. Computational Methods for Multiagent Simulations of Travel Behavior. Presented at 10th International Conference on Travel Behaviour Research, Lucerne, 2003.
52. Bernhardt, K. L. S. Agent-Based Modeling in Transportation. In *Transportation Research Circular E-C113: Artificial Intelligence in Transportation: Information for Application*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 72–80.
53. Arentze, T., and H. Timmermans. Social Networks, Social Interactions, and Activity-Travel Behavior: A Framework for Microsimulation. *Environment and Planning B: Planning and Design*, Vol. 35, No. 6, 2008, pp. 1012–1027.
54. Stephan, C. H., M. Mahalik, T. Veselka, and G. Conzelmann. Modeling the Transition to a Hydrogen-Based Personal Transportation System. Presented at Workshop of Frontiers in Transportation: Social Interactions, Amsterdam, 2007.
55. Sunitiyoso, Y., E. Avineri, and K. Chatterjee. On the Potential of Social Interaction and Social Learning in Modelling Travellers' Change of Behaviour Under Uncertainty. *Transportmetrica*, Vol. 7, No. 1, 2011, pp. 5–30.

Urban Travel Demand Forecasting

DAVID REINKE
Kittelson & Associates, Inc.

This article discusses applications of artificial intelligence (AI) methods to urban travel demand forecasting. The following article begins with an historical overview of urban travel demand forecasting.

HISTORICAL OVERVIEW OF URBAN TRAVEL DEMAND FORECASTING

Urban travel demand forecasting has its roots in the large-scale regional transportation planning that began in the 1950s. Planning efforts in Chicago, Detroit, and other areas developed what was to become known as the "four-step" travel demand forecasting process, where a region is divided into a number of transportation analysis zones and the models, as is also mentioned in the previous article by Avineri, estimated the following on a zonal basis (*1*, *2*).

- Trip generation: How many trips are being made?
- Trip distribution: From where to where are the trips going?
- Mode split: By which mode of travel (auto or transit)?
- Assignment: By which specific route?

The models in this process were known as aggregate models because they were based on aggregate characteristics of transportation analysis zones and they forecast aggregate behavior (e.g., total trips to or from a zone).

Transportation planning concerns began to change in the late 1960s and early 1970s and led to further model developments. In particular, the earlier models were strongly biased toward highway planning; they were difficult to apply to regional forecasting for alternatives that involved major changes to transit. The mode split step of the four-step process was especially criticized as being insensitive to changes in transit service.

Beginning in the late 1960s and 1970s a new approach to travel analysis was developed: the use of individual choice models of travel behavior. This approach was intended to predict the probabilities that an individual would choose a transportation mode. McFadden formulated this concept in terms of economic utility theory (*3*). Subsequent developments have extended and refined the formulation of choice models to cover a variety of ways of representing travel choices. Train gives a well-balanced presentation of the current state of the art in choice modeling (*4*). Choice models were eventually extended to include not only mode choice, but choice of destination (the trip distribution step) and trip generation as well. One of the earliest regional travel models entailing extensive use of disaggregate choice models was developed for the San Francisco Bay Area in the early 1980s (*5*).

The newer models still, however, had serious limitations. For example, non-home–based trips (trips that neither begin nor end at home) were particularly difficult to model because they

depended on the complete daily travel pattern of the individual. This led to efforts to develop models of an individual's complete daily pattern, the so-called tour-based models; at this time it was also desired to capture the nature of travel as a derived demand that depended on other activities an individual engages in, which led to the development of activity-based models. The distinction between activity-based and tour-based models is somewhat artificial, as some activity-based models have in fact been implemented as tour-based models. The term "activity based" will be used from here on. An extensive set of papers on activity-based modeling can be found at the U.S. Department of Transportation Travel Model Improvement Program website (*6*).

The 1990s saw the introduction of inexpensive computing and advances in statistical techniques that allowed the development of highly complex models of travel behavior. The movement began toward the development of activity-based or tour-based models, where the complete daily travel pattern of an individual would be modeled. Early examples include activity-based models developed for a road pricing study in Portland (*7*) for the City and County of San Francisco (*8*), and, more recently, for the San Francisco Bay Area (*9*).

Logit models have become increasingly prominent in urban travel demand. Formerly used for mode choice only, they are now used in activity-based models to model all choices. For example, in the Portland model (*7*), the choice hierarchy is as follows:

- Primary activity (mandatory, maintenance, or discretionary; either on tour or at home);
- Time of day (when the tour begins and ends); and
- Simultaneous choice of mode and primary destination.

Each of these choices is modeled using a logit model, where the higher-level models incorporate log sum utilities from the lower-level models in the choice utility function.

But logit models have some significant disadvantages for choice modeling. For example, unimodality of the log likelihood function is guaranteed only if the model is linear in parameters. [See McFadden (*3*) for a proof of this.] This greatly restricts the functional form of the model and limits the extent to which interaction terms can be included. And logit models can be quite sensitive to errors in the data. If the likelihood function is flat at the optimum, small errors in the data can lead to large errors in the parameter estimates.

OPPORTUNITIES FOR AI IN URBAN TRAVEL DEMAND MODELING

This section discusses some potential AI applications for urban travel demand forecasting. The focus in this article is on the travel demand modeling process itself, as opposed to the focus on traveler behavioral modeling in the paper by Avineri. The discussion takes two perspectives: concepts perspective (i.e., how certain AI concepts could improve the current state of practice), and a methods perspective (i.e., which AI methods could be applied and how).

Concepts Perspective

Two AI concepts that especially to urban travel demand modeling are model testing and the bias-variance trade-off. Both concepts are intimately related.

The almost exclusive practice in developing urban travel demand models is to (*a*) select models that best fit the data and (*b*) once the models are estimated, adjust them so that outputs from the model for a given base year match observed counts. This has two serious flaws:

- It does not test the dynamic behavior of the model, i.e., how well it performs for different land use and transportation system inputs.
- By using goodness-of-fit to the estimation data as the only model selection criterion, models tend to become overfitted to the estimation data.

The second flaw, overfitting, is especially a problem for the newer activity-based models, which typically have many hundreds of parameters. For example, the Portland activity model (*7*) has more than 500 parameters; the San Francisco activity model (*8*) has more than 600 parameters; the new San Francisco Bay Area activity model (*9*) has more than 1,600 parameters. In general, there is a trade-off between goodness-of-fit of a model to the estimation data and its performance on new data. This is illustrated in Figure 1, which shows typical model performance on training and test data sets; as the model becomes more complex, it fits the training data better. Test error also decreases with increasing model complexity, but only up to a point; beyond that point, the model becomes overfitted to randomness in the training data and the test error, which measures generalization performance of the model, increases. Like regression models, the logit model can also be evaluated by a criterion that penalizes overfitting. The Bayesian information criterion (*10*) is typically reported by most logit model estimation software but this is reported for a single model only; there is as yet no generally accepted criterion that penalizes overfitting on a large nested set of choice models.

Hence, one significant application of AI concepts to urban travel modeling is a prescriptive application: model selection should be based on performance of test data, and not merely goodness-of-fit to the training data.

Perhaps the best characterization of the key difference between AI approaches (in the context of machine learning) and traditional model-building approaches is that the AI approach emphasizes predictive performance as the primary criterion for model selection. Vapnik presents a brief summary of the differences in model-building approaches between traditional statistical theory and machine learning (*11*).

AI Methods

Given the increasing importance of discrete choice modeling in urban travel demand modeling, the most important contribution of logit modeling would appear to be in the area of choice modeling. Several candidates for alternatives to the logit model are neural networks (NN), support vector machines (SVM), and the relatively new method of relevance vector machines (RVM).

High bias
Low variance

Low bias
High variance

Error

—Training

—Test

Model complexity
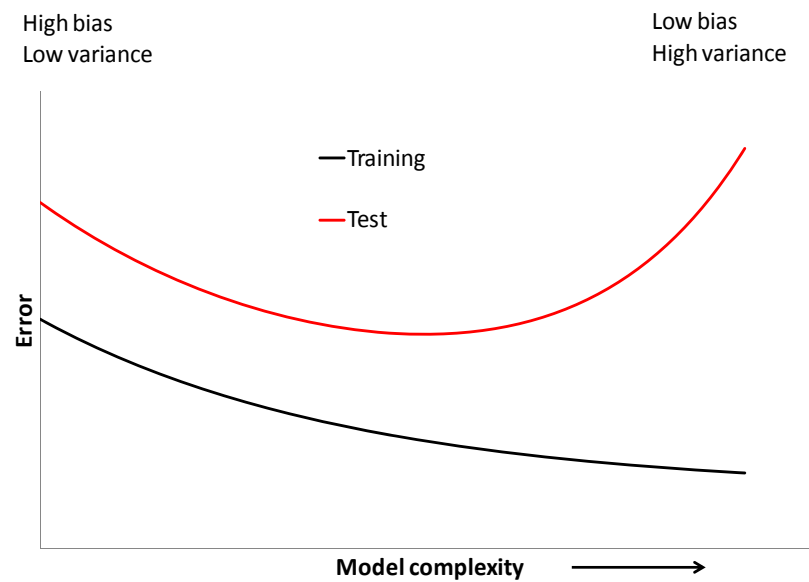
FIGURE 1 Illustration of bias–variance trade-off (*9*).

Neural Networks

Of the available AI methods, NNs appear most often in the research literature (*10, 12*). NNs have several advantages over traditional discrete-choice models, including the following:

- Nonparametric. NNs do not require a prespecified functional form in order to be fitted. The only requirement for the activation functions are that they be smooth and nonlinear.
- Nonlinearity. NNs can capture nonlinear effects much more readily than linear-in-parameters logit models.
- Interactions. NNs can also capture interaction terms between exogenous variables without having to specify interactions in advance.

But NNs have some significant limitations, including the following:

- Overfitting. Having too many hidden layers, neurons can lead to too many weight parameters, which in turn can lead to NNs that are seriously overfitted to the data. Reed discusses some of the methods used to overcome overfitting including pruning and weight decay (*12*). But these methods are somewhat ad hoc and rely extensively on the judgment of the modeler.
- Multiple local minima. The NN objective function can exhibit multiple local minima, making it difficult to converge on an optimal set of weights.

See Hastie et al. (*10*) and Bishop (*14*) for more complete discussions of neural networks and their uses.

Support Vector Machine

The SVM (*13–16*) is a relatively recent AI development that deals with some of the shortcomings of NNs. SVMs can be used for classification and regression problems. For classification (choice) problems, the SVM can be thought of a linear discriminant model that works in a high-dimensional space. The SVM approach to classification is as follows:

- Using a kernel function, transform the input data into a high-dimensional space. The transformation is typically nonlinear.
- In this high-dimensional space, find a hyperplane that provides the maximum separation between classes.

Practically, a classification SVM is estimated using a quadratic programming approach that optimizes the dual of the primal problem: minimizing the sum of squared weights plus a penalty term for misclassification.

The SVM has been shown to be superior to NNs for many applications (*13*). The SVM automatically calculates which training cases (the support vectors) determine the location of the separating hyperplane in high-dimensional space. Any positive-semidefinite kernel can be used; a radial basis function is a frequently used.

But SVMs are not without disadvantages:

- They are still susceptible to overfitting. See Bishop (*14*) for a discussion on how SVMs can overfit.
- Fitting a SVM requires specification of a C parameter that balances having a large separation margin versus a low misclassification error rate and another parameter for the kernel. These parameters are estimated by repeated training of the model and testing on a separate data set.
- The SVM is a "hard" classifier, where an input is assigned with certainty to a class. It therefore does not capture the probabilistic nature of choice represented in a logit model. Some types of SVM have been proposed that do allow for predicting probabilities (*9*).
- The SVM is difficult to apply when there are more than two classes. Several variations have been proposed, including simultaneous multiclass SVMs and nested SVMs, but each of these has its own disadvantages (*15*).

Relevance Vector Machines

The RVM is a relatively new development that was first proposed by Tipping (*17*). The regression form of the model works as follows. Assume we have a regression problem that is linear in parameters:

$$y(\mathbf{x}) = \sum_{1 \le i \le N} w_i \phi_i(\mathbf{x})$$

where $y(\mathbf{x})$ is the predictor, and $\phi_i(\mathbf{x})$ are a set of functions of the data $\mathbf{x}$, and w_i are the weights. In practice, the $\phi_i(\mathbf{x})$ are represented by a kernel function $\phi_i(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}_n)$; unlike the SVM, the kernel functions are not required to be positive semidefinite.

The RVM takes a Bayesian to estimating the weights w_i; each w_i is assumed to be have a prior Gaussian distribution:

$$p\left(w_i \mid \alpha_i\right) = N\left(w_i, 0, \mid \alpha_i^{-1}\right)$$

where the hyperparameters α_i represent the inverse variance, or precision, of the distribution. The posterior distribution of the w_i are then estimated by maximizing the likelihood function of the regression. This is done as an iterative process with repeated estimations of the w_i and α_i until convergence is reached. For a number of w_i, the hyperparameter α_i will be driven to infinity, indicating that that particular w_i is distributed around zero with zero variance; for those w_i, the associated kernel function drops out of the regression.

The classification form of the RVM works similarly to a logit model, only the utility function is replaced by a generalized regression model as shown above.

The RVM has some significant advantages over the SVM and NN:

- It typically has a sparser representation than the SVM and NN. Hence, it is computationally faster to apply.
- It is a Bayesian approach with a statistically consistent method for estimating the hyperparameters. Unlike the SVM, where the C and other parameters must be found by trial and error, the RVM automatically determines the appropriate weights as part of the optimization process.
- It is directly applicable to multiclass classification problems and avoids the difficulties encountered with the SVM classifier.

The main disadvantages of the RVM are that it is slower to train and that, for classification problems, the size of the training problem scales as the size of the training data set multiplied by the number of classes. However, these disadvantages are mostly offset by the faster application speed of the sparser RVM model.

CONCLUSIONS

This paper discussed the potential applications of AI to urban travel demand modeling. The emphasis has been on choice modeling, as this is the main component of the newer travel demand models. AI can contribute to travel demand modeling in two general ways.

- It provides a perspective on model selection that emphasizes the predictive capability of the model rather than its ability to fit a set of training data. In particular, overfitting is viewed as a significant problem that must be addressed if the predictive performance of the model is not to be compromised.
- It provides a set of techniques that can overcome some of the significant disadvantages of current modeling techniques used in practice.

AI methods are not without their drawbacks. While a number of readily available software packages are available for estimating traditional regression and logit models, there is

less available software for NNs, SVMs, and RVMs, but there is increasing availability of estimation software for the first two.

Another, perhaps more daunting, drawback of AI methods is that it is more difficult to interpret the results of training a model. Whereas a traditional logit model can provide easy-to-interpret parameters—such as the relative importance of waiting time and in-vehicle time in the mode choice decision, or the economic value of time—parameter estimates for AI methods cannot be as readily interpreted. A possible ways of addressing this latter problem would be the development of software to generate graphical response contours for combinations of variables that would make the model results more transparent.

REFERENCES

1. Oi, W. Y., and P. W. Shuldiner. *An Analysis of Urban Travel Demands.* Northwestern University Press, Evanston, Ill., 1962.
2. Lane, R., T. J. Powell, and P. P. Smith. *Analytical Transport Planning*. Duckworth, London, 1971.
3. McFadden, D. Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in Econometrics* (P. Zarembka, ed.), Academic Press, New York, 1974.
4. Train, K. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, 2009.
5. Cambridge Systematics, Inc. Travel Model Development Project: Phase 2 Final Report. Prepared for Metropolitan Transportation Commission, Oakland, Calif., 1980.
6. U.S. Department of Transportation. Travel Model Improvement Program. Available at http://tmip.fhwa.dot.gov/.
7. Cambridge Systematics, Inc. A System of Activity-Based Models for Portland, Oregon. DOT-T-99-02. U.S. Department of Transportation, 1998.
8. Cambridge Systematics, Inc. San Francisco Travel Demand Forecasting Model Development. Report prepared for San Francisco County Transportation Authority, 2002.
9. Metropolitan Transportation Commission. MTC Activity-Based Model System Development. Available at http://www.mtc.ca.gov/maps_and_data/datamart/abm/.
10. Hastie, T., R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2009.
11. Vapnik, V. N. *The Nature of Statistical Learning Theory*. Springer, New York, 2000.
12. Reed, R. D. *Neural Smithing*. MIT Press, Cambridge, Mass., 1998.
13. Schölkopf, B., and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, Mass., 2002.
14. Bishop, C. M. *Pattern Recognition and Machine Learning*. Springer, New York, 2006.
15. Abe, S. *Support Vector Machines for Pattern Classification*. Springer, London, 2005.
16. *Advances in Kernel Methods: Support Vector Learning* (B. Schölkopf, C. J. C. Burges, and A. J. Smola, eds.), MIT Press, Cambridge, Mass., 1999.
17. Tipping, M. E. Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*, Vol. 1, 2001, pp. 211–244. Available at http://jmlr.csail.mit.edu/.

Transportation Safety Analysis

NICOLAS SAUNIER
Ecole Polytechnique de Montréal

PRAVEEN EDARA
University of Missouri-Columbia

HOJONG BAIK
Korea Aerospace University

Transportation by any mode involves some risk of harm as described in the *Handbook of Transportation Science* (*1*): "the primary goal of transportation, the effective movement of people and goods, is better served by ever increasing speeds." The issue is that "in general, as speed increases so does risk." A vehicle striking anything is referred to as a collision. In this work, safety is defined as the number of collisions expected to occur at a given location per unit of time, where expected means "the average in the long run if it were possible to freeze all prevailing conditions that affect safety" (*2*).

The purpose of studying transportation safety is to examine the many factors that influence the likelihood of occurrence of, and the resulting harm or losses from collisions. A difference in one of those factors could have led to a different outcome. Hakkert and Braimaster define risk as "a combination of the probability, or frequency, of occurrence of a defined hazard and the magnitude of the consequences of the occurrence" (*3*). The consequences of collisions include fatalities, various levels of injuries and property damage. Interventions to improve safety may target the two components of risk:

- Collision prevention refers to measures aimed at preventing the collision from occurring.
- Crashworthiness refers to engineering features aimed at reducing losses, given that a specific collision occurs.

Addressing the uncertainty in safety data is a challenge. Even if there were no problems of data availability and quality, it is difficult to attribute fatalities and injuries to specific modes, as reporting requirements and definitions are not standardized across modes and jurisdictions. Depending on the transportation mode, the risk of collision has different characteristics. For example, most of the risk in air travel is related to taking off and landing, and therefore does not increase with the distance traveled, while it does for road travel. The risk depends mostly on the pilot for air travel whereas in case of road travel, risk depends on the individual road driver. Most of this section will be devoted to road safety since it has the largest mode share in the overwhelming majority of countries.

ROAD SAFETY

Road collisions involve various types of vehicles, motorized and nonmotorized, including bicycles and animal-powered vehicles as well as pedestrians and fixed objects; vehicles and pedestrians will be referred to as road users. There has been a considerable amount of research to estimate safety models as a function of explanatory variables describing the transportation system. The explanatory variables can be viewed as falling typically into one of the following three categories: the road, the vehicle, and the driver. A safety model, also called a crash prediction model (CPM) or safety performance function, typically takes the form of an equation linking safety to a set of variables.

Safety may be considered along the time and space dimensions. Comparing safety for different periods allows study of temporal trends and evaluation of the impact of countermeasures in studies before and after the implementation of the counter measure, called before-and-after studies. Comparing safety for different locations or spatial entities is done to identify locations that are particularly hazardous and thus have the most promising potential for safety improvements, which is done through rankings and prioritization.

The traditional and most common way of studying safety relies on collision databases that provide large amounts of data, much more than for any other mode. Collisions attributes describe the driver (age, gender), the road (location characteristics, weather) and the vehicle (make, model).

The source of these databases comes from police and insurance reports, filled manually after the collision; the collision is reconstituted rather than observed. These data are therefore limited and biased in many ways. For example, property damage-only types of collisions are underreported and reports are skewed towards the attribution of responsibility, not the search for the causes that led to a collision. Some extra details such as the vehicle speeds before the collision may be estimated through in-depth reconstitution analysis, which are costly and therefore not routinely performed. Even through in-depth reconstitution, the series of events that led to a collision may be only guessed at. An in-depth analysis requires the actual observation of the whole event up to the collision, including information about the trajectories for all the road users, and in particular the ones involved in the collision, from which speeds, accelerations, distances and other variables may be derived.

Yet a fundamental issue is that such safety evaluations are reactive; collisions have to occur to evaluate safety. Research has been undertaken to develop proactive methods for safety analysis relying on surrogate measures of safety. The most famous and utilized methods rely on traffic conflicts (near-misses), other traffic events, and less-severe road users' interactions, studied in a safety perspective.

SAFETY ANALYSIS AND ARTIFICIAL INTELLIGENCE

As mentioned above, safety models link a set of variables describing the road system to safety and other collision characteristics, such as collision outcomes. Historically, these models were developed based on collision records stored in databases. More recently, there has also been an interest in attempting to develop collision models and perform safety analysis based on actual observations of collisions and interactions as recorded, for example, through video and various in-vehicle sensors, as done in naturalistic driving experiments. In either case, artificial

intelligence (AI) and advanced computing techniques are particularly suited to mine the data, find associations and learn models from the data. The following sections will briefly discuss the application of AI methods to safety data analysis. Separate sections will be dedicated to discussing the methods based on collision records and the methods based on actual observations of collisions and other safety-related events.

Safety Analysis Based on Collision Records

Markov Chain Monte Carlo Method Within Fully Bayesian Inference

Classical parameter estimation techniques include maximum likelihood estimation, ordinary least square method, and others. As an alternative to classical parameter estimation techniques, the fully Bayesian (FB) inference techniques have become popular in traffic safety research. In essence, the classical techniques offer point estimates of parameters, but Bayesian inference methods generate distributions of the parameters. The main motivation of the transition to Bayesian method is to have parameter estimation tools applicable to generalized collision-frequency models without any limitation in underlying assumptions in the model. Another motivation of the transition is to minimize the bias resulting from the so-called regression-to-the-mean phenomena (*4, 5*).

In modeling collision frequency within FB inference scheme, a Poisson–Gamma model structure is widely adopted. In the model, the collision counts is assumed to be Poisson distributed with an average collision rate, and the prior distribution of the average collision rate is assumed to be a function of explanatory variables with the error term assumed to be Gamma distributed (*6*). From a computational perspective, however, posterior distributions of parameters are often intractable mainly because of the high dimensionality or the unavailability of standard form of distributions involved. This practical difficulty can be readily overcome by the Markov chain Monte Carlo (MCMC) simulation method that draws from the posterior distributions, in a univariate manner, assuming the other parameters to be fixed in each drawing. There are two popular types of MCMC methods: Gibbs sampler and Metropolis–Hastings (MH) algorithm. Gibbs sampler is preferable when the posterior distribution can be expressed as a standard form, whereas MH algorithm is a solution when the posterior distribution is in a nonstandardized form (*7*). There are several computational packages developed for Bayesian inference using the MCMC method. For example, WinBUGS developed by Cambridge University is publicly available, and SAS statistical package also provides MCMC procedure for Bayesian inference (*8*).

It should be noted that the simple Poisson model assumes that mean and variance of collision data are the same. This homogenous assumption does not fit the data well if the data are heterogeneous (i.e., over-dispersed or under dispersed). In an effort to better characterize the dispersed data, various types of statistical models have been suggested and tested. The models vary depending on how to assume the distribution of error terms and how to handle the crash data skewed toward zero occurrences. Among others, popular models tested include binominal, Poisson, Poisson–Gamma, Poisson–lognormal, Poisson–inverse Gaussian, zero-inflated Poisson, and zero-inflated negative binomial, etc. (*6, 9*).

In a similar context, Qin et al. set up a hierarchical Bayesian model using a zero-inflated Poisson model with a nonlinear model that represents the relationship between collision count and traffic volume (*10*). The parameters were estimated using MH algorithm for collision counts

and traffic volume data on rural two-lane highway segments collected from four different states in the United States. The model also considered other physical characteristics such as the daily volume, segment length, speed limit, and lane or shoulder width as explanatory variables. The results indicated that the relationship between collisions and the daily volume is highly nonlinear, and it also showed that there is no significant difference in the collision exposure functions from state to state.

Ma and Kockelman pointed out that most of previous studies relied on univariate modeling approach where collision counts are modeled separately by severity type (*11*). To address this shortcoming, they presented a multivariate Poisson (MVP) model in which the collisions and severity are jointly modeled. The model was developed in a Bayesian inference paradigm, and parameters were estimated using Gibbs sampler and MH algorithm for collision data collected on Washington State rural highways. Later, Ma et al. proposed a MVP-lognormal (MVPLN) model using Bayesian inference (*12*). The model was also estimated using a Gibbs sampler and MH algorithm for the same Washington State rural highway data. The resulting model shows that there exists a statistically significant correlation between collision counts by severity type, and also demonstrates that MVPLN can address over-dispersed data effectively. The paper also provides practical recommendations: wide lanes and shoulders are critical factors to reduce collision frequencies.

In an effort to develop a more generalized collision frequency model that can handle over- and under-dispersed collision data simultaneously, Lord et al. and Geedipally et al. adopted a Bayesian Conway–Maxwell–Poisson (COM-Poisson) generalized linear model (*13*, *14*). COM-Poisson model is a generalized Poisson model with an extra parameter for the rate of decay of successive ratios of parameters. Several versions of COM-Poisson models were set up in a hierarchical Bayesian framework, and evaluated using MCMC method for hypothetical data. The paper concluded that COM-Poisson models can flexibly handle over-, equi-, and under-dispersed collision data, and the MCMC procedure results in accurate parameter estimation.

To improve the accuracy of the crash prediction model, Lord and Miranda–Moreno examined the effects of low mean values and small sample size on the posterior mean of dispersion parameter involved in the Poisson–Gamma distribution (*15*). A hierarchical Bayesian model is set up and estimated by MCMC method using hypothetical low mean and small sample scenarios. It was concluded that low mean and small sample size seriously affect the estimation of the posterior mean of the dispersion parameter, which suggested that, in this case, Poisson–lognormal model is a better option than Poisson–Gamma model. A finite mixed Poisson and negative binomial models were also proposed and evaluated using MH algorithm for collision data collected at 868 signalized intersections in Toronto, Canada (*16*).

In addition to the collision-frequency model, MCMC method in association with Bayesian inference was also applied to rank collision-prone locations with high collision occurrence (*17–21*).

Machine Learning Methods

Several machine learning methods have also been applied to analyze collision data and to identify factors contributing to collisions. Among supervised methods, many popular forms of artificial neural networks (ANNs) have been explored: back propagation neural networks (BPNNs), Bayesian neural networks (BNNs), ANNs with radial basis functions (RBFs), and

neuro-fuzzy methods, as well as newer support vector machines (SVM). Well-known unsupervised methods have also been tried, among which hierarchical clustering and k-means.

CPMs have been developed using various models. Awad and Janson estimated crash prediction models for truck collisions at freeway ramps in Washington state (*22*). The performance of two types of AI models: neural network model, hybrid neural network, and fuzzy logic model were compared with linear regression model. The dependent variable was the truck collision frequency at each ramp location and the independent variables consisted of both qualitative and quantitative variables. The variation in the data was captured predominantly by 10 independent variables: gore-to-taper distance, ramp length, ramp average daily traffic (ADT), main road ADT, ramp truck ADT, main road truck ADT, weather conditions, visibility, road surface conditions, and location (urban or rural). The authors found that both types of AI models produced better fits compared to the regression model for training data. However, they report that the AI models over-fitted the training data and therefore were not able to satisfactorily predict the testing data.

Xie et al. developed CPMs for rural frontage roads in Texas using three different methods: (*a*) BPNN, (*b*) negative binomial (NB), and (*c*) BNN. Explanatory variables included segment length, ADT, right shoulder width, and lane width (*23*). The output variable was number of collisions in 5 years occurring at the site. It was shown that the BNN model outperformed the other two models. The authors point out that the data over-fitting is not an issue with BNN models, unlike BPNN models, and hence have better generalization capabilities. Building upon the previous work, Li et al. developed crash prediction models using SVM and found that the SVM models produced more accurate predictions than the NB models and produced the same or better results than the BPNN models (*24*). They also report that the SVM models are faster to implement than BPNN models.

Mussone et al. employed ANN to analyze vehicular collisions in Italy. A three-layer neural network model was proposed to estimate a collision index (defined as the ratio of the number of collisions at an intersection to the number of collisions in the most dangerous intersection) of urban intersections (*25*). Multiple linear and exponential regressions were tried but not pursued given their performance. The results showed that the ANN model can identify factors that contribute to intersection collisions: the degree of danger is the highest in nighttime collisions, for any type of collision, at signalized intersections, and for the running over of pedestrians at nighttime at nonsignalized intersections with respect to all types of collisions.

Kumara et al. used the generalized regression neural network in conjunction with subtractive clustering methodology in fuzzy logic to model the hazardousness of intersection approaches (*26*). If an intersection approach exhibits a higher potential for collisions than an ordinary location it is considered to be hazardous. Road collisions in Singapore were analyzed in the study. The dependent variable is the annual injury collision frequency of the intersection approach and the independent variables included traffic variables (approach volume, turning movement volumes), geometric data (sight distance, gradient, etc), and other variables such as presence of surveillance cameras, median railing, number of signal phases, and the presence of permissive right-turn phase. The prediction model was found to perform satisfactorily and also to work well in multidimensional measurement spaces where approach hazardousness is believed to be a function of several geometric and traffic measures.

Safety models for some collision characteristics have been also built to classify collisions and identify collision factors. Sayed et al. proposed a method to identify collision-prone locations based on a classification of the factors (related to the road, the driver or the vehicle)

that contribute to collisions (*27*). A fuzzy k-means algorithm is used to classify collisions using 14 description variables. The output is the degree of membership of each collision to the three categories of factors. The approach was validated on three datasets of 300 collisions, and compared to expert assessment. In a follow-up study, Sayed and Abdelwahab investigate the classification of road collisions using ANN and fuzzy k-nearest neighbours (*28*). Each collision was described by 21 variables, and the output was the same three membership degrees to road, driver, or vehicle factor. The performance of both models is considered good by the authors, with a slight advantage to ANN.

Roberts et al. present an experimental comparison of two inductive logic programming (ILP) algorithms (PROGOL and TILDE) with a propositional learning algorithm for decision trees (C4.5) on a propositional dataset of 1,413 reports of road traffic collisions that occurred in Great Britain in 1996 (*29*). The goal was to predict whether the collision was caused by a young driver. Rebalancing methods were used to handle the skewed distribution of positive and negative examples in this dataset, and the different relative cost of errors of commission and omission. Before rebalancing, all algorithms perform worse than majority class and improved significantly after. The authors conclude that on such a dataset, ILP algorithms perform competitively in terms of predictive accuracy with propositional systems, but are significantly outperformed in terms of training time taken (by a factor of 100 to 600).

Several attempts at modeling collision outcomes (i.e., injury, property damage) have also been made. Sohn and Shin tried various models, namely neural networks, decision trees, and logistic regression, to identify collision severity-related factors using a database of 11,564 traffic collisions collected in 1996 in Seoul, Korea (*30*). The output was binary: bodily injury (death or injury) and property damage. They found that the presence of a protective device (i.e., seatbelt or helmet) is the most important factor in the cash severity variation. Other factors include collision type, speed before collision, violent driving, road width, and car shape. The classification accuracy of these three techniques was also compared without finding any significant differences among them. Judging this classification accuracy to be relatively low, Sohn and Lee propose various algorithms to improve the accuracy of the individual classifiers, neural networks, and decision trees (*31*). Three different approaches were tried: (*a*) classifier fusion based on the Dempster–Shafer method, the Bayesian procedure, and logistic model; (*b*) data ensemble fusion based on arcing and bagging; and (*c*) clustering based on the k-means algorithm. Their empirical study results indicate that a clustering-based classification algorithm works best for their data.

Multilayer perceptron (MLP) and RBF neural networks were developed to analyze the traffic safety at toll plazas in Florida by Abdelwahab and Abdel-Aty (*32*). Collision location and injury severity probabilities were estimated using MLP, RBF, and logit models. The data included all variables related to collision, driver, vehicle, toll plaza, road conditions, and environment conditions. The study found that the RBF neural network model was the best model for analyzing driver injury severity and the nested logit model slightly outperformed the RBF model for estimating collision locations. Abel-Aty and Abdelwahab investigated the use of two well-known ANN paradigms, MLP and fuzzy adaptive resonance theory (ART) neural networks, to analyze driver injury severity given that a crash has occurred (*33*). Modeling results showed that the testing classification accuracy was 73.5% for the MLP and 70.6% for the fuzzy ARTMAP, compared to 61.7% for a calibrated ordered probit model. The results of the models showed that gender, vehicle speed, seat belt use, type of vehicle, point of impact, and area type (rural versus urban) affect the likelihood of injury severity levels.

Chong et al. summarizes the performance of four machine learning paradigms applied to modeling the severity of injury resulting from traffic collisions: the models were neural networks trained using hybrid learning approaches, SVMs, decision trees, and a hybrid decision tree–neural network, a neural network with an extra input provided by the prediction of a decision tree on the same data (*34*). Using 10,247 records of the National Automotive Sampling System General Estimates System, the task was to classify the severity of injuries in five categories: no injury, possible injury, nonincapacitating injury, incapacitating injury, and fatal injury. Each class was learned against the others, and the results compared separately. Experiment results show that the hybrid decision tree-neural network approach outperformed the individual approaches.

Delen et al used ANN to determine important predictors of injury severity in roadway traffic collisions (*35*). The nonlinear relationships between injury severity levels and collision factors such as demographics of the person, environmental, and roadway conditions at the time of collision occurrence, vehicle characteristics (a total of 17 explanatory variables) are modeled using ANN. Five levels of injury severity—no injury (0), possible injury (1), nonincapacitating injury (2), incapacitating injury (3), and fatality (4)—were defined. A comprehensive dataset of 30,358 collision reports representative of a national sample was used to build and validate the models. The study findings were consistent with the previous studies on injury severity and provided new insights into the relationship between injury severity and contributing collision factors.

Chang and Wang developed a classification and regression tree to analyze the risk factors that can influence the injury severity in traffic collisions (*36*). The decision tree predicts the discrete output of no-injury, injury, and fatality based on 20 predictor variables that include temporal, highway–environmental, driver–vehicle characteristics, and collision variables. The model is learnt and tested on a data set of 26,831 road users involved in the collisions that occurred in the Taipei area in 2001. The results indicate that the most important variable associated with collision severity is the vehicle type. Pedestrians, motorcycle, and bicycle riders are identified to have higher risks of being injured than other types of vehicle drivers in traffic collisions.

Depaire et al used latent class clustering technique to segment traffic collision data to identify homogenous collision types (*37*). Injury models were then developed using multinomial logit model for each resulting cluster. Collision data from Belgium consisting of 29 variables describing collision, vehicle, road user, and environmental characteristics was used in the study. Seven clusters were formed based on features such as collision type, crossroad type, built-up area, road type, road user age, dynamics of road user (moving or stationary), and vehicle type. The study found that cluster models (*a*) revealed new variables affecting injury outcome and (*b*) provided a more complete interpretation of the relationship causal variables and injury outcome.

Shen et al. proposed to develop a unique road safety index from seven risk indicators to evaluate the overall safety performance and rank countries according to the number of road fatalities per million inhabitants (*38*). The proposed method relies on neural networks and rough sets that can be used for feature selection. The neural network learning algorithm assigns weights to the risk indicators, and rough sets identify speed as the most important attribute for the set of 21 European countries using the 2003 numbers. The new system ranking outperforms simple neural networks, but does not match so well the actual ranking.

ANNs have been used to develop an expert system for accident appraisal by Chiou (*39*). An accident appraisal is performed by a third party to examine the liabilities of all parties

involved in the accident. The expert system is trained with historical appraisal cases and emulates the behavior of appraisers. Benefits of such a system include shorter appraisal times, inheritance of appraiser's knowledge and experience, and consistent appraisal outcomes. In the study, appraisal liability is the output variable and is grouped into five categories with numerical values assigned to each group. The five groups are (1) full responsibility, (2) major responsibility, (3) equal responsibility, (4) minor responsibility, and (5) no responsibility. A total of 39 explanatory variables including the collision characteristics (date, time location, daylight, etc.), driver demographics (age, gender, etc.), traffic violations (speeding, alcohol use, etc.), driver behavior (direction, movement, etc.), and evidence (collision spot, driver injury, etc.) were used in the expert system.

Natural Language Processing

When collision reports have not been codified yet into databases, there is a need for natural language processing (NLP) techniques to extract relevant information automatically. Wu and Heydecker introduced a natural language understanding system to retrieve information from road collision reports in plain English (*40*). The grammar and lexicon were built on a database of 10,000 collisions in the United Kingdom. The results of the system are compared to the coded description of the collision, which is then confirmed or extended (with 1% of contradictions). More recently, Kayser and Nouioua presented a complete automated natural language understanding system that can determine the causes of a collision from a textual description of it (*41*). The system analyses the text description of a set of 160 collision reports from an insurance company (only self-contained texts were kept, not the ones relying on accompanying drawings). The analysis is made difficult because the collision report authors attempt to lessen their responsibilities in the collision, and because of the ambiguity of determining the cause or causes of collisions. This work demonstrates the feasibility of a robust and quick NLP system for collision reports, and highlights many areas for improvement.

Safety Analysis Based on the Observations of Collisions and Interactions

Traffic event observations are mostly obtained through video and various in-vehicle sensors, as done in naturalistic driving experiments. This section will focus on AI-based methods which can be used for safety analysis in this case.

Computer Vision

Different factors have led to a growing interest in the use of video sensors to collect safety data: falling prices for video sensors and computer hardware coupled with an increase in computational power, and the development of computer vision techniques that allow for the automated interpretation of video data. Readers are referred to Yilmaz et al. for a detailed overview of moving object tracking in video data (*42*).

　　　　The difficulty of tracking all road users depends on the complexity of the traffic scene. It is still an open problem in open outdoor intersections. If there is significant mixed traffic, the type of road users should be identified for safety analysis. Kamijo et al. process traffic images at intersections using a spatio-temporal Markov random field; the algorithm models a tracking problem by determining the state of each pixel in an image and its transit, and how such states

transit along both the *x–y* image axes as well as the time axes (*43*). The method is sufficiently robust to segment and track vehicles at a high success rate of 93% to 96%. Messelodi et al. mix region-based tracking, based on background subtraction, and feature tracking for efficient tracking (*44*). Classification is performed using predefined 3-D models of the road users. Hu et al. successfully use 3-D model-based vehicle tracking, but report some problems with occlusion (*45*). Saunier and Sayed, Saunier et al., and Ismail et al. rely on feature-based tracking since it does not require any particular initialization and it can deal with partial occlusion; the tracking accuracy for motor vehicles was measured between 84.7% and 94.4% on three different sets of sequences (*46–51*). Ismail et al. proposes an original road user classification method based on road users' typical motion patterns (*51*).

Machine Learning

A video analysis tool will output the road users' trajectories and sometimes their contour positions as well. AI techniques are then typically used for semantic interpretation. Among machine learning techniques, dynamic Bayesian networks, e.g., hidden Markov models (HMM) and Kalman filters, are particularly suitable to model sequential data such as trajectories and other continuous measurements of a traffic scene.

Messelodi and Modena describe an alternative approach based on occupation rates to evaluate the average probability of an accident for a passing vehicle. However, there is no direct detection of actual events (*52*). Collision probability is computed even for isolated vehicles without any other interacting vehicle in the intersection. Ismail et al. implement traditional traffic conflict indicators to study pedestrian–vehicle interactions (*50, 51*).

The system presented by Kamijo et al. learns various events and behavior patterns of each vehicle in the HMM chains and then, using the output from the tracking system, identifies current event chains (*43*). The system is tested on three collisions and can recognize bumping, passing, and jamming and may be extended to recognize other events, e.g., illegal U-turns or reckless driving by including these event patterns in the training set. Saunier and Sayed use HMM to learn the trajectories of road users in an unsupervised way; using a few traffic conflict instances, some clusters are adapted to represent the trajectories of road users involved (*46*). Detection is performed by matching observed road user trajectories to the clusters, the interactions of which trajectories are assigned to conflicting clusters are identified as traffic conflicts. The system is demonstrated on a small set of traffic conflicts.

Hu et al. use a fuzzy self-organizing neural network algorithm to learn activity patterns from the sample trajectories (*45*). Collisions are predicted by locating and matching each partial trajectory with the learned activity patterns, and the occurrence probability of a traffic accident is determined. Experiments show the effectiveness of the proposed algorithms on two accident instances in a toy car experience.

Saunier et al. propose an original incremental algorithm to learn motion patterns: trajectories are compared using the longest common subsequence similarity to identify prototype trajectories that represent each pattern (*47*). Adapting the formulas of Hu et al., the observed road users' trajectories are matched to the prototypes to predict their future positions and compute at each instant the probability of colliding at a later point (*45*). An advantage with respect to previous work and traditional traffic conflict techniques is the ability to study the severity of all traffic interactions to derive a more robust relationship to safety (*48*). This framework was refined by Saunier et al. by identifying potential collision points, displaying their spatial

distribution and computing the expected time to collision. It was demonstrated on a new dataset containing more than 200 traffic conflicts and 100 collisions (*49*).

Similarly, research has been undertaken to analyze data streams online; for example, for intersection monitoring, although most applications are vehicle based. Salim et al. propose techniques to discover crash patterns and precrash behavior associations, called ubiquitous data mining, based on a layered multiagent system (*53*). Horovitz at al. describe some algorithms tested on drunk driving detection using artificial data (*54*). Ning et al. proposed a framework for detecting unsafe system states using real-time traffic data obtained from sensors (*55*). They use temporal difference learning algorithm to train a danger-level function that can be utilized to inform drivers of hazardous situations on roadways. Interactive driving simulator experiments were performed to illustrate the framework.

CONCLUSION

Many AI methods have been applied to better understand the road safety problem, identify the various factors that favor collisions and severe outcomes, and develop more effective countermeasures. Although road safety has improved over the last decades in developed countries, it is an ongoing and growing health burden in developing countries which has lead the World Health Organization to launch in 2011 its Decade of Action for Road Safety.

AI and advanced computing techniques are part of the solution to reduce the harm done by transportation accidents. Particularly promising technologies are autonomous vehicles and the commoditization of sensors, means of communication, and computing power. Autonomous vehicles promise safe travel during which time can be spent usefully. An ever-growing and connected network of sensors will provide the so-called "big data" that can be interpreted by AI techniques and help us better understand the failures of the transportation system.

REFERENCES

1. Evans, L. *Handbook of Transportation Science*, Springer, New York, 2003, pp. 67–112.
2. Hauer, E., J. C. N. Ng, and J. Lovell. Estimation of Safety at Signalized Intersections. In *Transportation Research Record 1185*, TRB, National Research Council, Washington, D.C., 1988, pp. 48–61.
3. Hakkert, A., and L. Braimaister. *The Uses of Exposure and Risk in Road Safety Studies*. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid, Netherlands, 2002.
4. Hauer, E. Observational Before–After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety. Pergamon, 1997.
5. Hauer, E., D. W. Harwood, F. M. Council, and M. S. Griffith. Estimating Safety by the Empirical Bayes Method: A Tutorial. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1784*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 126–131.
6. Lord, D., and F. Mannering. The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research Part A*, Vol. 4, No. 5, 2010, pp. 291–305.
7. Train, K. *Discrete Choice Methods with Simulation*. Cambridge University Press, Mass., 2009.
8. Chen, F. Bayesian Modeling Using the MCMC Procedure. SAS, 2009.

9. Lord, D., S. P. Washington, and J. N. Ivan. Poisson, Poisson-Gamma and Zero-Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accident Analysis and Prevention*, Vol. 37, No. 1, 2005, pp. 35–46.

10. Qin, X., J. N. Ivan, N. Ravishanker, and J. Liu. Hierarchical Bayesian Estimation of Safety Performance Functions for Two-Lane Highways Using Markov Chain Monte Carlo Modeling. *Journal of Transportation Engineering*, Vol. 13, No. 5, 2005, pp. 345–351.

11. Ma, J., and K. M. Kockelman. Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1950*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 24–34.

12. Ma, J., K. Kockelman, and P. Damien. A Multivariate Poisson-Lognormal Regression Model for Prediction of Crash Counts by Severity, Using Bayesian Methods. *Accident Analysis and Prevention*, Vol. 40, No. 3, 2008, pp. 964–975.

13. Lord, D., S. D. Guikema, and S. R. Geedipally. Application of the Conway-Maxwell-Poisson Generalized Linear Model for Analyzing Motor Vehicle Crashes. *Accident analysis and Prevention*, Vol. 40, No. 3, 2008, pp. 1123–1134.

14. Geedipally, S. R., S. D. Guikema, S. S. Dhavala, and D. Lord. Characterizing the Performance of the Bayesian Conway-Maxwell Poisson Generalized Linear Model. Technical report. Texas A&M University, 2008.

15. Lord, D., and L. F. Miranda–Moreno. Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter of Poisson-Gamma Models: A Bayesian Perspective. *Safety Science*, Vol. 46, 2008, pp. 751–770.

16. Park, B. J., and D. Lord. Application of Finite Mixture Models for Vehicle Crash Data Analysis. *Accident Analysis and Prevention*, Vol. 41, No. 4, 2009, pp. 683–691.

17. Miaou, S., J. J. Song, and B. K. Mallick. Roadway Traffic Crash Mapping: A Space-Time Modeling Approach. *Journal of Transportation and Statistics*, Vol. 6. No. 1, 2003, pp. 33–57.

18. Miaou, S.-P., and J. J. Song. Bayesian Ranking of Sites for Engineering Safety Improvements: Decision Parameter, Treatability Concept, Statistical Criterion, snd Spatial Dependence. *Accident Analysis and Prevention*, Vol. 37, No. 4, 2005, pp. 591–806.

19. Brijs, T., D. Karlis, F. Bossche, and G. Wets. A Bayesian Model for Ranking Hazardous Road Sites. *Journal of the Royal Statistical Society Series A*, Vol. 107, 2007, pp, 1–17.

20. Miranda–Moreno, L. F., A. Labbe, and L, Fu. Bayesian Multiple Testing Procedures for Hotspot Identification. *Accident Analysis and Prevention*, Vol. 39, No. 6, 2007, pp. 1192–1201.

21. Miranda–Moreno, L. F., L. Fu, S. V. Ukkusuri, and D. Lord. How to Incorporate Accident Severity and Vehicle Occupancy into the Hotspot Identification Process? In *Transportation Research Record: Journal of the Transportation Research Board, No. 2102*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 53–60.

22. Awad, W. H., and B. N. Janson. Prediction Models for Truck Accidents at Freeway Ramps in Washington State Using Regression and Artificial Intelligence Techniques. In *Transportation Research Record 1635*, TRB, National Research Council, Washington, D.C., 1998, pp. 30–36.

23. Xie, Y., D. Lord, and Y. Zhang. Predicting Motor Vehicle Collisions Using Bayesian Neural Network Models: An Empirical Analysis. *Accident Analysis and Prevention*, Vol. 39, No. 5, 2007, pp. 922–933.

24. Li, X., D. Lord, Y. Zhang, and Y. Xie. Predicting Motor Vehicle Crashes Using Support Vector Machine Models. *Accident Analysis and Prevention*, Vol. 40, No. 4, 2008, pp. 1611–1618.

25. Mussone, L., A. Ferrari, and M. Oneta. An Analysis of Urban Collisions Using an Artificial Intelligence Model. *Accident Analysis and Prevention*, Vol. 31, No. 6, 1999, pp. 705–718.

26. Kumara, S. S. P., H. C. Chin, and W. M. S. B. Weerakoon. Identification of Accident Causal Factors and Prediction of Hazardousness of Intersection Approaches. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1840*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 116–122.

27. Sayed, T., W. Abdelwahab, and F. Navin. Identifying Accident-Prone Locations Using Fuzzy Pattern Recognition. *Journal of Transportation Engineering*, Vol. 121, No. 4, 1995, pp. 352–358.
28. Sayed, T., and W. Abdelwahab. Comparison of Fuzzy and Neural Classifiers for Road Accidents Analysis. *Journal of Computing in Civil Engineering*, Vol. 12, No. 1, 1998, pp. 42–47.
29. Roberts, S., W. van Laer, N. Jacobs, S. Muggleton, and J. Broughton. A Comparison of ILP and Propositional Systems on Propositional Traffic Data. *Inductive Logic Programming*, Vol. 1446, 1998, pp. 291–299.
30. Sohn, S. Y., and H. Shin. Pattern Recognition for Road Traffic Accident Severity in Korea. *Ergonomics*, Vol. 44, No. 1, 2001, pp. 107–117.
31. Sohn, S. Y., and S. H. Lee. Data Fusion, Ensemble and Clustering to Improve the Classification Accuracy for the Severity of Road Traffic Accidents in Korea. *Safety Science*, Vol. 41, No. 1, 2003, pp. 1–14.
32. Abdelwahab, H. T., and M. A. Abdel-Aty. Artificial Neural Networks and Logit Models for Traffic Safety Analysis of Toll Plazas. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1784*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 115–125.
33. Abdel-Aty, M. A., and H. T. Abdelwahab. Predicting Injury Severity Levels in Traffic Crashes: A Modeling Comparison. *Journal of Transportation Engineering*, Vol. 130, No. 2, 2004, pp. 204–210.
34. Chong, M., A. Abraham, and M. Paprzycki. Traffic Accident Analysis Using Machine Learning Paradigms. *Informatica*, Vol. 29, No. 1, 2005, pp. 89–98.
35. Delen, D., R. Sharda, and M. Bessonov. Identifying Significant Predictors of Injury Severity in Traffic Accidents Using a Series of Artificial Neural Networks. *Accident Analysis and Prevention*, Vol. 38, 2006, pp. 434–444.
36. Chang, L.-Y., and H.-W. Wang Analysis of Traffic Injury Severity: An Application of Non-Parametric Classification Tree Techniques. *Accident Analysis and Prevention*, Vol. 38, No. 5, 2006, pp. 1019–1027.
37. Depaire, B., G. Wets, and K. Vanhoof. Traffic Accident Segmentation by Means of Latent Class Clustering. *Accident Analysis and Prevention*, Vol. 40, 2008, pp. 1257–1266.
38. Shen, Y., T. Li, E. Hermans, D. Ruan, G. Wets, K. Vanhoof, and T. Brijs. A Hybrid System of Neural Networks and Rough Sets for Road Safety Performance Indicators. *Soft Computing—A Fusion of Foundations, Methodologies and Applications*, 2009.
39. Chiou, Y.-C. An Artificial Neural Network-Based Expert System for the Appraisal of Two-Car Accidents. *Accident Analysis and Prevention*, Vol. 38, 2006, pp. 777–785.
40. Wu, J., and B. G. Heydecker. Natural Language Understanding in Road Accident Data Analysis. *Advances in Engineering Software*, Vol. 29, No. 7–9, 1998, pp. 599–610.
41. Kayser, D., and F. Nouioua. From the Textual Description of an Accident to Its Causes. *Artificial Intelligence*, Vol. 173, No. 12–13, 2009, pp. 1154–1193.
42. Yilmaz, A., O. Javed, and M. Shah. Object Tracking: A Survey. *ACM Computing Surveys*, Vol. 38, No. 4, 2006.
43. Kamijo, S., Y. Matsushita, K. Ikeuchi, and M. Sakauchi. Traffic Monitoring and Accident Detection at Intersections. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 2, No. 2, 2000, pp. 108–118.
44. Messelodi, S., C. M. Modena, and M. Zanin. A Computer Vision System for the Detection and Classification of Vehicles at Urban Road Intersections. *Pattern Analysis and Applications*, Vol. 8, 2005, pp. 17–31.
45. Hu, W., X. Xiao, D. Xie, T. Tan, and S. Maybank. Traffic Accident Prediction Using 3D Model Based Vehicle Tracking. *IEEE Transactions on Vehicular Technology*, Vol. 53, No. 3, 2004, pp. 677–694.
46. Saunier, N., and T. A. Sayed. Automated Analysis of Road Safety with Video Data. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2019*, 2007, pp. 57–64.

47. Saunier, N., T. Sayed, and C. Lim. Probabilistic Collision Prediction for Vision-Based Automated Road Safety Analysis. *Proc., 10th International IEEE Conference on Intelligent Transportation Systems*, Seattle, 2007, pp. 872–878.
48. Saunier, N., and T. A. Sayed. Probabilistic Framework for Automated Analysis of Exposure to Road Collisions. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2083*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 96–104.
49. Saunier, N., T. Sayed, and K. Ismail. Large-Scale Automated Analysis of Vehicle Interactions and Collisions. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2147*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 42–50.
50. Ismail, K., T. Sayed, N. Saunier, and C. Lim. Automated Analysis of Pedestrian-Vehicle Conflicts Using Video Data. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2140*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 44–54.
51. Ismail, K., T. Sayed, and N. Saunier. Automated Analysis of Pedestrian–vehicle Conflicts: Context For Before-and-after Studies. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2198*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 52–64.
52. Messelodi, S., and C. M. Modena. A Computer Vision System for Traffic Accident Risk Measurement: A Case Study Advances in Transportation Studies. *International Journal, ITC*, Vol. 7(B), 2005, pp. 51–66.
53. Salim, F. D., S. Krishnaswamy, S. W. Loke, and A. Rakotonirainy. Context-Aware Ubiquitous Data Mining Based Agent Model for Intersection Safety. *Embedded and Ubiquitous Computing*, Vol. 3823, Springer, 2005, pp. 61–70.
54. Horovitz, O., S. Krishnaswamy, and M. M. Gaber. A Fuzzy Approach for Interpretation of Ubiquitous Data Stream Clustering and Its Application in Road Safety. *Intelligent Data Analysis*, Vol. 11, No. 1, 2007, pp. 89–108.
55. Ning, H., W. Xu, Y. Zhou, Y. Gong, and T. S. Huang. A General Framework to Detect Unsafe System States from Multisensory Data Stream. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 11, No. 1, 2010, pp. 4–15.

Transportation Security

MASHRUR CHOWDHURY
Clemson University

RYAN FRIES
Southern Illinois University, Edwardsville

RICHARD BROOKS
Clemson University

Globalism is a significant factor in the current and future state of transportation security. In particular, Bloom argues that, as a result of globalization, the ability to find terrorists and their weapons has become a primary objective of the air transportation industry (*1*). Global terrorism networks are becoming increasingly amorphous and more difficult to track, particularly as members and their communication networks change in a seemingly random progression. Further, because information about available security technologies and techniques is readily available via the Internet, methods of beating these security tools can be developed anywhere. These factors indicate that regardless of the number or cost of security measures, there will always be an inherent risk of terrorism on transportation systems. The concern over terrorism is in spite of the minimal impact of terrorism on society as a whole. It has been mentioned that if terrorism were considered as a public health problem, the mortality rate is small enough that no funding would be given to combat it (*2*). Behavioral economics explains why the population is so concerned about this problem, in spite of it being a rare occurrence. We are loss adverse, especially when the potential loss is unbounded (*3, 4*).

As globalism continues to blend cultures throughout the world, perhaps those in industrialized nations will begin to accept that security risks can never be fully mitigated regardless of the forethought devoted or the amount of funding available. However, it is imperative that security risks to transportation system are minimized in an optimal fashion. Artificial intelligence (AI) and other advanced computing applications are potential tools for minimizing security risks in an efficient manner without the need for major investments in human resources. Questions remain as to whether we will optimize to objective standards or human perceptions of risk (*3, 4*).

Transportation infrastructure includes a vast network of physical interconnects as well as a less-visual network of sensors, communication systems, and software. For each of these systems, different factors influence the continued evolution of security efforts.

These factors include the continued advancement of computer processing capabilities, increasing data availability and quality, continued technology deployment, advancing globalism, and progressing sustainability initiatives. Despite all of this change surrounding the transportation security spectrum, there are multiple promising AI tools that can help transportation agencies and practitioners thrive instead of just survive in maximizing security apparatus in the multidimensional and multimodal transportation system.

APPLICATIONS OF AI IN TRANSPORTATION SECURITY AREAS

The following section presents a synopsis of these AI applications by different areas related to transportation. These security areas are the following:

- Disaster response and evacuation,
- Freight and commercial vehicle security,
- Maritime security,
- Hazardous material (hazman) security,
- Wide-area alert,
- Rail security,
- Transit security,
- Transportation infrastructure security, and
- Traveler security.

Disaster Response and Evacuation

These tools can utilize data from different devices (sensors, cameras, etc.) and agencies. The AI tools will use these inputs for detection and verification of disaster events. Automated detection and verification may use fewer resources and produce fewer errors than manual processing. Surveillance requires attentively processing hours of monotonous data streams in a search for rare anomalies. Where humans tend to become bored and inattentive, machines trained using AI techniques excel at reliably performing rote tasks. This use of AI can reduce errors, potentially saving lives, and possibly be less expensive.

Specific to evacuation, public transportation agencies develop and maintain an evacuation plan in case of an emergency. These evacuation plans include evacuation routes, traffic control plans, such as closures of roads and ramps, and operational strategies, such as contra-flowing traffic. These plans are developed based on anticipated traffic demands related to available evacuation time and the resources available to transportation agencies. As traffic conditions may change during actual evacuations, an AI-based decision-support tool can be utilized to help evacuation operations under a dynamic traffic condition.

Recent work has used agent-based traffic simulation to evaluate evacuation strategies on different road network types (i.e., street grids and ring roads) and population densities. The findings indicated that no strategy is best for all road network types and that evacuation time of high-density urban areas can be reduced by staging the evacuation orders (5). Although hurricanes are excellent applications for staged evacuations, events such as terrorist attacks require immediate evacuations and may not allow staging.

Other work has focused on the evacuating crowds rather than vehicles. Crowd dynamics has been frequently studied and crowds during an evacuation would exhibit the following four characteristics:

- Individuals follow lowest-cost paths at a constant walking speed (6).
- When people are sparse, interactions among individuals are negligible. In the aggregate, people move along the path of least resistance (lowest energy) and maintain a radius of open space around them (7, 8).

- As crowds become dense, the radii between individuals decrease and a friction factor hinders traffic flow and traffic lanes emerge. Interactions at intersections result in either alternating or circular traffic patterns (*9*).
- When crowd density becomes excessive, people become nervous and a number of pathological behaviors emerge. Individual behavior becomes erratic, but group behavior becomes more ordered and traffic flow slows. This slowing is referred to as "freezing by heating" (*10*). When crowd speed decreases below a certain threshold, panics can arise spontaneously (*11*).

To model these different characteristics, tools such as cellular automata, lattice gas, social force, and agent-based models have the ability to model the behavior of fleeing pedestrians microscopically. Due to the complexity of these models and the computational requirements, other studies have used macroscopic models such as fluid-dynamic models. Xiaoping et al. have noted a trend toward integrating more than one of these micro- or macroscopic models to model the behavior of evacuating pedestrians. Examples of these integrated approaches include cellular automata models based on social force and lattice gases (*12*).

Cellular automata have shown themselves particularly adept at predicting these phase changes. Unfortunately, the analysis of cellular automata has not produced satisfying analytical results. To date, we can mainly detect the conditions under which aggregate behaviors emerge (*13*).

This field of study has yet to address the diversity of evacuees or particularly physiological or psychological factors. Integrating existing models, especially those with human factors, shows promise of increasing the accuracy of pedestrian evacuation models (*6*).

Freight and Commercial Vehicle Security

In recent years, due to security and economic reasons, fleet and freight management companies have increased their monitoring of fleets and freight (*14*). Security applications leverage knowledge of what is being tracked (driver, fleet, or cargo) to identify security breaches. As with disasters, AI tools can potentially help identify deviations from the norm.

Often the conditions classified as the norm are found using Monte Carlo methods. The complexity of Monte Carlo analysis may be reduced by embedding it into a game tree analysis of the problem. Tree search can simplify the situations that need to be considered. This combination of tools is called Markov chain Monte Carlo (MCMC) analysis and can estimate pay-offs of these intermediate expressions. This idea has driven advances in computer algorithms for competing in the game Go, where MCMC finds likely future positions of the Go board. These heuristics that drive a probabilistic depth first tree search. Based on this approach, computer programs have been able to beat top-level Go players for the first time (*15*). Integration of game theory into security analysis is an important tool that shows much promise.

Maritime Security

A vast amount of the world's international trade uses maritime shipping links. Individually inspecting each container is well beyond the scope and budget of any country's security efforts. Instead, efforts are focused towards identifying high-risk containers and inspecting those. Because of the mountains of paperwork that accompanies these vast numbers of incoming

containers to every port (24 h prior to arrival), AI can play a significant role in assisting security officials. Artificial neural networks (ANNs) have been used to review the transport documents (bill of lading, consignment note, and receipt) to identify containers with higher probabilities of illegal items such as stolen merchandise. One ANN application used port of origin, shipping company, receiving company, type of goods shipped, and missing or incomplete shipping documentation to learn which items increased the historical security risks. This method was successfully validated and is awaiting implementation at the Port of Long Beach in California (*16*). Similar applications have the potential to provide much-needed focus to maritime port inspections.

Recent maritime piracy increases have identified a need for additional security efforts to prevent en-route disruptions to international cargo trade. There are consideration costs associated with security contractors, insurance, and fuel for rerouting vessels. Additionally, there is a plethora of open sea shipping route choices and security countermeasures from which shippers can choose. AI tools are particularly appropriate for helping to identify the best options. One study used an agent-based simulation to identify optimum countermeasures and routes for a particular trip (*17*). Although there is no guarantee of preventing piracy, these AI tools can assist shipping companies in better risk management and preparations for each shipment.

Hazardous Material Security

Hazmat security entails minimizing the risk of (*a*) hijacking of security-sensitive hazmat cargo and (*b*) use of hazmat cargo as a weapon. Security applications in this area will include tracking of hazmat-carrying vehicles, roadside detection of hazmat cargo, and authentication of drivers of commercial vehicles, especially those carrying hazmats. AI tools have the potential to assist with route selection during the planning phase and to assist in detection of security breaches while shipments are enroute. As onboard Global Positioning Systems continue to flood the trucking market, AI algorithms can assist transportation managers to track key vehicles to ensure route adherence and identify potential security breaches.

Wide-Area Alert

This application covers notifying traveling public in emergency situations, such as disaster warnings, and civil emergencies. AI-based decision-support tools could be utilized to identify the severity of an impending disaster and generate appropriate notifications for public. These systems can supplement any systems currently used in traffic management centers for traffic incident management, tornado warnings, or hurricane evacuations. The string of tornados across the U.S. Midwest and South in spring 2011 emphasizes the continued need for timely and accurate emergency notification to the public by all methods available.

Rail Security

This area includes securing trains and rail related assets and personnel. One of the applications of rail security is highway–rail intersection. Security application will include surveillance and operation of the highway–rail intersection. AI can be used in surveillance to detect any potential threat. Various commercial surveillance systems can include AI algorithms to identify the presence of stopped vehicle on a rail crossing or those entering restricted areas.

AI could be utilized in developing a decision-support system to manage if an incident has occurred at the highway–rail intersection. For example, future AI systems could prompt emergency dispatchers to inform the appropriate agencies and personnel and could provide cues to the dispatcher, ensuring the right type of information is provided to each notified agency. Such a tool may be instrumental to the future of managing all incidents by assisting clear communication among responders.

Transit Security

Transit security includes security of passenger, assets, and facilities for passenger rail and bus transit systems. Video surveillance is an important tool to monitor transit stations, vehicles, and other transit-related assets. AI tools can be utilized to develop useful information received from multiple cameras.

Although much current security information is gathered and analyzed at a centralized location, Remagnino et al. evaluated a security system where a distributed AI system analyzed video from multiple cameras. This study found that AI was an appropriate tool for identifying and classifying events based on video data (*18*). Other studies have reported advantages of decentralized or distributed intelligence for incident detection for their scalability and efficiency (*18–20*). It is thus appropriate that distributed AI systems be considered for future transit security applications, in addition to application in other types of incident detection system in the transportation network.

Critical Transportation Infrastructure Security

Transportation infrastructure includes bridges, tunnels, multimodal facilities, and transportation management centers. AI-based algorithms with their learning capabilities allow sensor and surveillance systems to improve in detections and predictions over time, as well as adapt to changing conditions. ANN has been the most commonly applied AI tool for traffic incident detection. Another AI paradigm called support vector regression (SVR) and support vector machines (SVM) have also been applied to detect incidents. The underlying theory behind SVM and SVR is similar. SVM is primarily used for pattern classification, whereas SVR is used for regression or function estimation. So far, SVR has had limited applications within the transportation field. Previous examples of applying SVR to transportation-related problems include its use for travel time, traffic speed, and traffic flow predictions, as well as for incident detection (*21–24*).

Traveler Security

Traveler security entails safety and security of travelers in public facilities, including at transit stations, parking facilities, and transportation infrastructure. ANN has also been applied to facial recognition for security applications. Facial recognition algorithms have demonstrated a detection rate as high 90% (*25*) and show promise for tracking suspicious individuals in and around transportation facilities with high-quality surveillance systems. As that quality improves and the cost of video surveillance decreases, there is a potential to identify and track wanted criminals as they travel our public transportation systems, improving law enforcement abilities.

Privacy concerns will likely play a significant role in the future of public policy on the security of transportation hubs particularly with video recording.

CONCLUSIONS

AI tools can support transportation security areas ranging from disaster response and evacuation to a traveler personal security. The advantages of using AI will be cost savings, error reduction, and increased reliability due to automating security systems. Although no approach can ever mitigate the risk of all threats, AI tools can improve the efficiency of the limited funds dedicated to transportation security each year.

REFERENCES

1. Bloom, R. W. Fear of Flying: Globalization, Security, and Terrorism. *TR News*, No. 269, July–August 2010, pp. 21-27.
2. Anderson, R. Available at http://www.tech.dmu.ac.uk/STRL/news/annual-seminar/STRL-ADS-2009.m4v. Accessed May 2011.
3. Gintis, H. *The Bounds of Reason*, Princeton University Press, 2009.
4. Ariely, D. *Predictably Irrational*, Harper Perennial, 2010.
5. Chen, X., and F. B. Zhan. Agent-Based Modeling and Simulation of Urban Evacuation: Relative Effectiveness of Simultaneous and Staged Evacuation Strategies. *Journal of the Operational Research Society*, Vol. 59, 2008, pp. 25–33.
6. Hoogendoorn, S. P., and P. H. L. Bovy. Pedestrian Route Choice and Activity Scheduling Theory and Models. Delft University, Netherlands, 2002.
7. Weidmann, U. Transporttechnik der Fussgaenger, Institut fuer Verkehrsplannung, Transporttechnik, Strassen- und Eisenbahnbau, ETH Zuerich, 1993.
8. Helbing, D. Traffic and Related Self-Driven Many-Particle Systems. Available at Arxiv:cond-mat/0012229v2. Accessed April 23, 2001.
9. Molnar, P. Modellierung un Simulation der Dynamik von Fussgaengerstroemen. Shaker, Aachen, 1996.
10. Helbing, D., I. Farkas, and T. Vicsek. *Physical Review Letters*, Vol. 84, 200, p. 1240.
11. Helbing, D., I. Farkas, and T. Vicsek. *Nature*, Vol. 407, 2000, p. 487.
12. Xiaoping, Z., Z. Tingkuan, and L. Mengting. Modeling Crowd Evacuation of a Building Based on Seven Methodological Approaches. *Building and Environment*, Vol. 44, 2009, pp. 437–445.
13. Wolfram, S. *A New Kind of Science*, Wolfram Media, Springer, New York, 2002, pp.1–22.
14. Fries, R., M. Chowdhury, and J. Brummond. *Transportation Infrastructure Security Utilizing Intelligent Transportation Systems*. John Wiley and Sons, Hoboken, N.J., 2008.
15. Coulom, R. Efficient Selectivity and Backup Operators in Monte Carlo Tree Search. *Proc., 5th International Conference on Computers and Games*, Turin, Italy, 2006.
16. Kiang, M. Y. Cargo Security Early Warning System: The Application of Neural Networks to Detect Cargoes with Potential Security Fraud. Final Report. METRANS Project 07318906, 2008. Available at http://www.dot.ca.gov/hq/research/researchreports/reports/2008/ar_06-03_final_report.pdf .
17. Jakob, M., O. Vanek, S. Urban, P. Tenda, and M. Pechoucek. AgentC: Agent-Based Testbed for Adversarial Modeling and Reasoning in the Maritime Domain. *Proc., 9th International Conference on Autonomous Agents and Multiagent Systems* (van der Hoek, Kaminka, Lesperance, Luck, and Sen, eds.), Toronto, Canada, 2010, pp. 1641–1642.

18. Remagnino, P., A. I. Shihab, and G. A. Jones. Distributed Intelligence for Multi-Camera Visual Surveillance. *Pattern Recognition*, Vol. 37, 2004, pp. 675–689.
19. Chowdhury, M., and K. C. Wang. Transport Science and Technology. *Distributed Intelligent Traffic Sensor Network*, Elsevier, Amsterdam, Netherlands, 2007.
20. Ma, Y., M. Chowdhury, A. Sadek, and M. Jeihani. Real-Time Highway Traffic Condition Assessment Framework Using Vehicle-Infrastructure Integration with Artificial Intelligence. *Proc., IEEE Transactions on Intelligent Transportation Systems*, Vol. 10, No. 4, 2009, pp. 615–627.
21. Cheu, R. L., D. Srinivasan, and E. T. Teh. Support Vector Machine Models for Freeway Incident Detection. *Proc., Intelligent Transportation System*, Vol. 1, 2003, pp. 238–243.
22. Ding, A., X. Zhao, and L. Jiao. Traffic Flow Time Series Prediction Based on Statistics Learning Theory. *Proc., IEEE 5th International Conference on Intelligent Transportation System*, 2002, pp. 727–730.
23. Vanajakshi, L., and L. R. Rilett. A Comparison of the Performance of Artificial Neural Networks and Support Vector Machines for the Prediction of Traffic Speed. *Intelligent Vehicles Symposium,* IEEE, 2004, pp. 194–199.
24. Wu, C. H., J. M. Ho, and D. T. Lee. Travel-Time Prediction with Support Vector Regression. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, No. 4, 2004, pp. 276–281.
25. Rowley, H. A., S. Baluja, and T. Kanade. Neural Network-Based Face Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, 1998, pp. 23–38.

Analysis of Public Transportation Planning and Operations

SHINYA KIKUCHI
Virginia Tech

Public transportation transports people in aggregated form. Aggregation takes place in space and time. Spatial aggregation occurs at the stations (stops) and in the network, while time aggregation is accommodated by the predetermined departure and arrival schedule. By aggregating passengers in a limited number of vehicles and corridors, public transportation offers many benefits over private auto transportation. It reduces traffic congestion by decreasing the number of vehicles on the roads, it promotes energy efficiency, and it also promotes compact and mixed land use. In order to realize these benefits of public transportation, however, the parameters of transit operations and services must be selected carefully. The parameters are many and they are interrelated in a complex manner. The processes for planning, designing, and operating transit services involve various trade-offs among the selection of the parameter values.

Analysis of transit operation has traditionally been conducted within the classical framework of optimization using the statistical approach or classical mathematical optimization. This section reviews the opportunities which AI and advanced computing can offer in order to overcome some of the analytical barriers in the traditional approaches.

ANALYSIS OF TRANSIT PLANNING AND OPERATIONS

Transit analysis typically deals with the determination of parameter values at different stages: planning, design, operations–management, and policy making. At the planning stage, the issues include demand projection, modal integration, land use along the network, environmental and energy effects, and the evaluation of alternatives. At the design stage, the focus is to determine the degree of passenger aggregation including headways, stop locations, the network, fleet size, and facilities. At the operations–management stage, the issues are crew assignment, fleet management, schedule adjustment and recovery, labor and equipment management, fare collection, and the collection of operating data. At the policy development stage, the issues include the effects on modal share, financing, pricing transportation, economic impacts, social equity, and the role of transit in the overall accessibility provision. The research opportunities for artificial intelligence (AI) and advanced computation are found in each of these stages, especially for the topics related to multiobjective optimization, knowledge acquisition, and data collection–management. Difficulties in transit analysis are found in three areas.

First, representing and applying available knowledge for inference. The knowledge that is available may be incomplete, fragmented, or dependent on language-based rules. The problem is how to preserve and represent such knowledge in a quantitative manner so that the analytical process can be carried out systematically. Neural network and fuzzy inference systems will have a part in the representation of this knowledge.

Second, how to manage the data about transit planning and operations is an important issue. Transit operation involves the handling of large amounts of statistical data and frequent

data updates. The agency must collect data on the ridership, revenue, incidences, operating performance, labor management, equipment and maintenance records, facility inventory and maintenance, and service(s) offered. Not only is it important to collect and maintain data, but is important to apply it and extract useful information from it. The data mining ability of AI must be exploited for this function. In other words, data are used to analyze the operating characteristics but they should also be used for discovery of system characteristics.

Third, the parametric analysis of transit operation involves optimization, or obtaining optimum combinations of parameter values. Optimization must satisfy the desires of individual stakeholders, rather than maximizing or minimizing a converted single objective function. The traditional optimization approach, in which all objectives are put together in a single function with weights to each, is no longer acceptable in today's decision-making environment in which many stakeholders' desires must be accounted for individually. AI has a role in representing the desires of individual stakeholders, and formulating the problem as a multiobjective optimization. Transit planning is indeed a multiobjective optimization process.

Finally, the critical aspect of transit analysis is to present the reasoning process to the public and to the decision makers and to show how much is known and how much is not known. This analysis involves measuring uncertainty and the interplay(s) between uncertainty and information.

AI AND TRANSIT ANALYSIS: MULTIOBJECTIVE ANALYSIS

Selection of parameter values must consider the interests of many stakeholders, e.g., the users, the transit agency, labor, the community, the local and regional governments that fund the transit services, and different population segments which are affected differently. AI methods can offer a participatory decision tool in which each stakeholder takes part in the process of parameter selection. This section presents an example that uses a fuzzy multiobjective optimization framework.

Multiobjective Optimization

The selection of transit system parameter values must be based on the satisfaction of individual stakeholders, rather than the optimization of a global single function. This attitude is a major departure from the traditional optimization concept. Further, the analysis process is not top-down, but bottom-up, in which each stakeholder's requirements and desires, as well as the global requirements, are defined first.

Consider the case where the values of a set of transit parameters, $x_1, x_2, x_3, \ldots, x_n$, must be determined in order to satisfy a set of objectives and constraints. Assume that there are three objectives to achieve: G1, G2, and G3. Each objective may be characterized initially by "maximize," "minimize," or "maintain at this value." In addition, there are several constraints which define the relationships among the decision parameters.

In many cases, the decision makers do not really mean "maximum" or "minimum" in the literal sense. "Maintain at this value" may actually mean maintain *around* or *approximately* this value. What is important is the sense of "satisfaction" in terms of achieving the individual objectives. In other words, the decision is made in the contextual world, not in the absolute world. Thus, a function which represents the sense of satisfaction can be developed by each

stakeholder under a particular decision context. The exact shape of this function with respect to the achievement of objective may differ among the analysis situation and the stakeholder, case by case.

Figure 1 shows examples of the satisfaction function. Figure 1a shows the shapes of the membership function representing "approximately X," Figure 1b shows the shapes of "large" and "very large" relative to b. Figure 1c shows the case of "small" and "very small" relative to 0. Note that the specific shapes of these functions can be determined by the individual stakeholders, not by the single decision maker or analyst.

Let us assume the following situation, although abstract, in which the decision variables, $x_1, x_2, x_3,\ldots, x_n$, and the objectives are related:

$$\text{Objective 1: G1 should be small, where } G1(x_1, x_2, x_3,\ldots, x_n) \tag{1}$$

$$\text{Objective 2: G2 should be very small, where } G2(x_1, x_2, x_3,\ldots, x_n) \tag{2}$$

$$\text{Objective 3: G3 is approximately equal to A, where } G3(x_1, x_2, x_3,\ldots, x_n) \tag{3}$$

G1 may represent the satisfaction of a "small wait time"; G2 may represent a "very short access distance"; G3 may represent an "acceptable cost of operation." Each function represents the level of satisfaction of the respective objective for the set of values of x_i's.

Let us denote the functional relations among the decision parameters as $R_1, R_2,\ldots, R_m$, where $R_i (x_1, x_2, x_3,\ldots, x_n) = C_i$, $i = 1$ to m, and m = the total number of relationships. Some of these relations can be an inequality or a fuzzy relation. The optimization model that satisfies the three goals, G1, G2, and G3, and the relationships among the decision variables becomes the following:

$$\max h \tag{4}$$

$$G1(x_1, x_2, x_3,\ldots, x_n) \geq h \tag{5}$$

$$G2(x_1, x_2, x_3,\ldots, x_n) \geq h \tag{6}$$

$$G3 (x_1, x_2, x_3,\ldots, x_n) \geq h \tag{7}$$
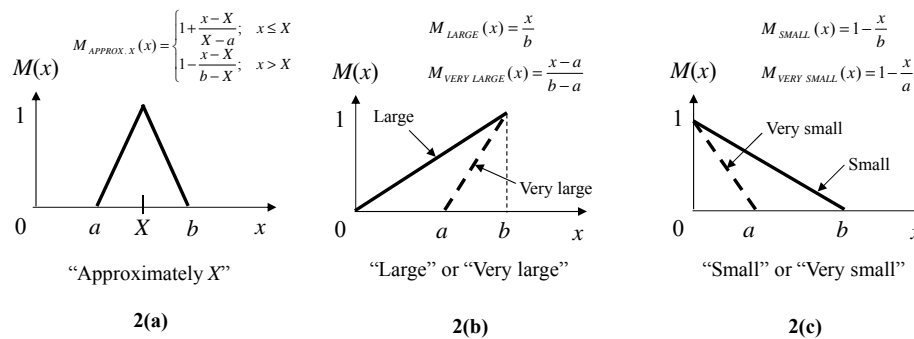
FIGURE 1 Examples of membership functions for representing satisfaction.

$$R_i(x_1, x_2, x_3,\ldots, x_n) = C_i, \; i = 1, \, m \tag{8}$$

This formulation allows the satisfaction level of the least-satisfied objective to be maximized, that is, the max–min concept. Note that $R_i(x_1, x_2, \, x_3,\ldots, x_n) = C_i$ acts also as an objective in this formulation. In other words, all constraints are in essence the objectives, because they must be satisfied also. This problem can be solved by linear programming if all constraints are in linear additive form; if not, it can be solved by a nonlinear optimization model. The obtained value of h represents the least satisfaction level; in other words, all objectives are satisfied at least at the value of the obtained h.

This concept was originally conceived by Bellman and Zadeh (*1*); later, Zimmermann (*2*) formulated it as linear programming problem and coined the term *fuzzy optimization* for this process; and Dubois and Prade (*3*) called it *egalitarian optimization*. Applying this concept, Kikuchi et al. (*4–7*) developed multiobjective optimization models to solve problems in transportation engineering including traffic signal timing and calibration of traffic counts.

EXAMPLE

Let us consider a simple example in which one wishes to determine the "optimum" headway between transit vehicles for a scheduling problem. The optimum headway is a compromise between the interests of the user and those of the operator. The shorter the headway, the shorter the user wait time becomes; at the same time, it results in a greater cost of operation due to the need for a greater fleet size as well as for associated expenses. Hence, the problem is to determine the headway that satisfies the user's and the operator's concerns.

In the classical method, the headway that minimizes total cost, a sum of the user cost and the operator cost, is considered the optimum headway. In this process, the user's wait time is converted to cost using the value of time. The objective becomes minimizing the total cost S, the sum of the user cost and the operator cost in the following equation:

$$\min S = w_1 \, U(h) + w_2 \, O(h) \tag{9}$$

where $U(h)$ and $O(h)$ are the costs per unit time for the user and the operator, respectively; and w_1 and w_2 are the weight assigned to the user cost and the operator cost, respectively. $U(h)$ and $O(h)$ are costs related to headway through the user wait time and the fleet size.

We can assume that the wait time is a function of headway, h:

$$U(h) = C_u \, x \, D(h) \, x \, P \tag{10}$$

where C_u is the user's value of time spent waiting for the vehicle (\$/time), and $D(h)$ is the user wait time when the headway is h. If the headway is very short (say, $h < 15$ min), $D(h)$ may be assumed to be ½ h. P is the passenger volume per unit time.

We can also assume that the operator's cost is a function of headway.

$$O(h) = C_o \, x \, F(h) \tag{11}$$

where C_o is the operator's cost (not just the cost of the vehicle but all associated costs) per vehicle per unit time. $F(h)$ is the fleet size when the headway is h for the given vehicle capacity and the characteristics of the line. For example, the relation can be $F(h) = C/h$, where C is the cycle time. In this case, the assumption is that all of the operator's costs are related to fleet size, which is not an unrealistic assumption since fleet size dictates the costs for operation, vehicle maintenance, and personnel or administration.

The value of h which minimizes the total cost S in Equation 9 is found by differentiating it with respect to h (headway). However, this formulation brings about the following problems: first, the determination of the values of weights, w_1 and w_2; second, the meaning of "minimum S"; and third, the consistency of the measurement units.

First, w_1 and w_2 are the weights in the objective function. Yet, their values are hard to determine. They seem to show the relative weight of the user's dollar with respect to the operator's dollar. The realistic values for w_1 and w_2 are impossible to determine in a meaningful manner.

Second, the meaning of S, which is to be minimized, is not clear. Some say it is the total social cost, but it is not known if this is the true interest of the decision maker or of society. Each concerned party is interested in minimizing its own cost, rather than an elusive total cost. Hence, the acceptability of the solution by each party is not taken into account.

Third, the user cost in this problem is usually based on the wait time cost experienced by the hourly or daily passengers whereas the operator's cost is most likely based on the annual cost. The timeframe of the cost calculation is not consistent. It is possible that depending on the timeframe, the optimum value of h may be different.

The proposed AI approach can circumvent some of these problems, using the formulation shown in Equations 1 through 8. It separates the interests of the user and the operator, and maximizes the satisfaction of these parties, separately. Thus, the problem becomes a multiobjective optimization problem. The following shows the formulation:

$$\max q \tag{12}$$

$$S_{user}(T) \geq q; \; S_{operator}(C_o) \geq q \tag{13}$$

$$T = D(h); \; C_o = F(h) \tag{14}$$

where $S_{user}(T)$ and $S_{operator}(C_o)$ are the satisfaction functions of the user and the operator, respectively.

These functions represent the satisfaction of the user with respect to the wait time (T), and also of the operator with respect to the total cost (C_o). For the user, the smaller the value of T is the better; for the operator as well, the smaller the value of C_o, is the better.

Because the desires of the user as well as the operator are both of the nature of "the smaller, the better," the shapes of the membership functions, $S_{user}(T)$ and $S_{operator}(C_o)$, are of the type shown in Figure 1c. The specific shape of the function depends on the expression of the user under the specific circumstances. It may be that the user is satisfied as long as the wait time is less than or near 5 min. For the operator, a similar situation may occur; as long as the cost is less than a given value, it is acceptable. In the classical model, the objective was the elusive concept of the *minimum*. Furthermore, having each concerned party draw their satisfaction functions

rather than having the third party decide on the objective of optimization is a more realistic and natural representation of desire.

The formulation shown in Equations 12 through 14 maximizes the least satisfied party's satisfaction. Thus, each party's concern is represented. The satisfactions of the two parties are determined independently, and no consideration is given to the consistency of the timeframe in setting up the cost function as seen in the classical model. In other words, the time unit of the user cost can be the hourly cost, and the operator cost can be the annual cost. Further, there is no need to consider the relative weights for the user and the operator.

It is interesting to see that the strength of satisfaction or desire of the user and the operator can be represented not using the relative weights w_1 and w_2, as in the classical model. Representation of the strength of desire can be handled by two ways: one, by adjusting the shape of the satisfaction function, and two, by adding constraints that represent the priorities among the objectives in the formulation. For the former, if the user's desire for a short wait time is "very" strong, then a shape of the satisfaction function that is similar to "very small" in Figure 1c can be chosen. Similarly, the satisfaction function for the operator can be changed by the desire for the operator. For the latter, if the user's satisfaction is more important than that of the operator's, the problem can be reformulated as follows:

$$\max q_2 \tag{15}$$

$$S_{\text{user}}(T) \geq q_1; \ S_{\text{operator}}(C_\text{o}) \geq q_2, \tag{16}$$

$$T = D(h); \ C_\text{o} = F(h), \tag{17}$$

$$q_1 \geq q_2 \tag{18}$$

where $q_1 \geq q_2$ assures the level of satisfaction of the users is greater than that of the operator.

SUMMARY

How to represent the desire and satisfaction of individual stakeholders is an important requirement when modeling a problem in which many stakeholders' interests are intertwined. This is usually the case in the analysis of the parameters of public transportation. What is shown in this chapter is an example of multiobjective optimization in which the objectives are expressed by the satisfaction level of the individual stakeholders. The concept and formulation is relatively straightforward and easy to apply to real-world problems.

One of the most useful aspects of this approach is the representation of the desires of the individual stakeholders by the satisfaction functions. The shape of the satisfaction function can be freely drawn by the stakeholder first, and later, it can be adjusted to a form that can be approximately represented by the mathematical function.

When many parameters are involved in a complex manner, the principle of decomposition and aggregation should be applied. First, decompose the problem into smaller parts and clarify the parametric relationships within each part and between parts, and then aggregate the parts to form the whole. This is precisely the approach taken in the above example; first, the interest of the user and the operator are separated and their desires are represented by

the satisfaction functions; second, these are aggregated by the max–min operation to obtain the values of decision parameters. Thus, the least-satisfied party's satisfaction is maximized, the egalitarian concept. Further, this multiobjective formulation considers the integrity of the desires of the individual stakeholders. The traditional mathematical formulation expresses the desire of all stakeholders by a single function. Two elements of AI in this formulation are the use of the satisfaction function by the membership function of fuzzy set, and the bottom-up approach of separating the interests of individual stakeholders.

It is interesting to consider the connection between the proposed approach and agent-based system. Agent-based system is characterized by the independent behavior of individual agents each pursuing to maximize its utility. A complex system may be considered as a collection of autonomous decision-making entities and it assesses its situation and acts according to a set of rules that maximizes its utility. For the definition of agent-based system reader may refer to North and Macal (*8*), Srbljinovic (*9*), and Bonabeau (*10*).

What is discussed in this paper is, in fact, optimization of an agent-based system in which the system, as a whole, reaches the balance in terms of satisfying the desires of every agent. While agent-based system is a representation of the many agents interacting and the process of searching for the overall balance of desire. Optimization of a system as it was presented here is a mathematical approach to reach the balance. The parameters of public transportation are actually the agents. Each parameter represents the interest of some stakeholders, such as the users, the operator, the labor, the community, etc. The formulation presented provides an overall balance of satisfaction where the individual stakeholder can specify its degree of satisfaction.

To date, little work has been seen in the application of AI to the analysis of public transportation. This is in sharp contrast to AI's application to highway traffic analysis, as the previous articles of this circular clearly demonstrate. A closer coordination between transit analysts and AI specialists is needed to identify the problems that may be conducive to the AI approach and identify the benefits of such effort. It is a responsibility of the TRB AI committee to take the initiative in approaching transit and planning agencies, and promoting the potential benefits of AI approaches to solve some of their problems.

REFERENCES

1. Bellman, R., and L. A. Zadeh. Decision-Making in a Fuzzy Environment. *Management Science*, Vol. 17, No. 4, 1970, pp. 141–164.
2. Zimmerman, H. J. *Fuzzy Set Theory and Its Applications* (4th edition). Kluwer, Boston, Mass., 2001.
3. Dubois, D., and H. Prade. Advances in the Egalitarist Approach to Decision-Making in a Fuzzy Environment. *Dynamical Aspects in Fuzzy Decision Making* (Y. Yoshida, ed.), Physica-Verlag, Heidelberg, 2001, pp. 213–240.
4. Kikuchi, S., S. Mangalpally, and A. Gupta. Method for Balancing Observed Boarding and Alighting Counts on a Transit Line. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1971*, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 42–50.
5. Kikuchi, S., and D. Miljkovic. Method to Preprocess Observed Traffic Data for Consistency: Application of Fuzzy Optimization Concept. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1679*, TRB, National Research Council, Washington, D.C., 1999, pp. 73–80.

6. Kikuchi, S., D. Miljkovic, and H. J. van Zuylen. Examination of Methods that Adjust Observed Traffic Volumes on a Network. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1717*, TRB, National Research Council, Washington, D.C., 2000, pp. 109–119.
7. Kikuchi, S., and Y. S. Murat. Fuzzy Optimization Approach: Comparison with the Classical Optimization Method Using the Problem of Timing a Traffic Signal. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2024*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 82–91.
8. North, M. J., and C. M. Macal. Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modelling and Simulation. Oxford University Press, New York, 2007.
9. Srbljinovic, A., and O. Skunca. Agent Based Modelling and Simulation of Social Processes. *Interdisciplinary Description of Complex Systems*, Vol. 1, 2003, pp.1–8.
10. Bonabeau, E. Agent-Based Modelling: Methods and Techniques for Simulating Human Systems. In *Proceedings of National Academy of Sciences*, Vol. 99, 2002, pp.7280–7287.

Design and Construction of Transportation Infrastructure

M. G. KARLAFTIS
National Technical University of Athens

S. M. EASA
Ryerson University

M. K. JHA
Morgan State University

E. I. VLAHOGIANNI
National Technical University of Athens

Maintaining, restoring, and improving urban infrastructure has been recognized as one of the engineering grand challenges of the 21st century by the National Academy of Engineering. The challenges arise from both the design and the construction of such infrastructure. Designing a new highway is a very complex problem that should combine data from various sources such as design specifications, costs, safety, geographic features, and even environmental and political impacts of the alignments (*1*). Moreover, transportation infrastructure is usually designed not only to address the demand for transportation but also to benefit socioeconomic growth via trip time reduction, safety improvement, noise and air pollution reduction, and so on. Cost-intensive transportation infrastructure design, in many ways, involves trade-offs that are largely governed by the stated goal of the targeted application, such as safety, serviceability, and accessibility. These goals should be jointly considered along with environmental and economic issues, land use restrictions, and so on. Achieving each of these objectives might require different approaches that directly affect the parameterization of the design considerations, leading to more complex and difficult-to-solve design problems.

As for transportation infrastructure construction, complexities arise from the distinct characteristics of each type of infrastructure. Highways, airports, bridges, bus terminals, canals, docks, railroads, roads, train terminals, and tunnels have different requirements in terms of planning, scheduling, and cost estimation. Moreover, transportation infrastructure is strongly related to urban development patterns; the impact of the infrastructure systems to urban growth is an enduring and "unsolved mystery" (*2*). Funderburg et al. showed that, while improvements in surface transportation tend to have large impacts on growth patterns, the nature of the effects is materially dependent on the context of the highway investment (*2*).

In such a complex environment, the decisions that should be made are critical and continuously changing. Sundin and Braban–Ledoux stress that, as transportation systems become increasingly complex, the number of available alternatives increases (*3*). In addition, the cost of making errors can be very large because of the complexity and magnitude of operations and automation; further, the chain reaction that an error can cause in many parts of the transportation system cannot be discounted.

Any decision regarding the design and construction of transportation infrastructure should be prompt and linked to the entire transportation system infrastructure. Transportation networks typically span a large geographical area and have a modular structure that consists of many subsystems with actors and sensors, while involving both continuous and discrete dynamics that evolve over different time scales (*4*). The difficulty in such a case arises from the fact that most often the available information is quite intensive; in a network-level thinking, a large amount of transportation assets should be taken into consideration that, many times, may impose conflicting design and construction goals (*5*).

Need for Artificial Intelligence

Nowadays, most transportation agencies continuously collect enormous streams of data from different sources for monitoring and controlling the transportation facilities; the task of fusing these data into a comprehensive dataset to be used for the design and construction of transportation infrastructure may prove exceedingly difficult for the usual classical statistical modeling. This is because of two issues. First, the multivariate nature of the data imposes statistical modeling structures that are difficult to develop, test, and understand. Second, frequently, real-time collected data incorporate both missing values and a significant amount of uncertainty requiring intelligent ways of analyzing them (*6*).

In particular, in transportation infrastructure design and construction problems, there is the need to jointly consider information that may be measured and quantified and other information that is subjective and should be qualitatively assessed, such as driver perception, comfort, safety and so on (*7*).

Artificial intelligence (AI) provides an alternative to facing the difficulties mentioned above by providing methodologies for developing flexible multivariate models for solving difficult approximation and optimization problems. The flexibility of AI methods [e.g., knowledge-based systems, expert systems, pattern recognition, machine learning, neural networks, genetic algorithms (GAs) and evolutionary computation, fuzzy systems, etc.), as well as their performance in various interdisciplinary applications are well suited for the complexity and variety of transportation systems.

AI has been applied to various fields of transportation engineering (for a review see *3*, *6*, *8*, *9*). AI has also played a critical role in solving many infrastructure problems related to design and construction (including maintenance). Some of the AI techniques that are most popular in infrastructure design and construction are GAs, simulated annealing, ant algorithms (AAs), and fuzzy logic (FL).

AI APPLICATIONS IN INFRASTRUCTURE DESIGN AND CONSTRUCTION

### Application Areas and Related AI Techniques

Table 1 summarizes the available literature of AI applications in infrastructure design with respect to the specific problem faced and the technique applied. As can be observed, the oldest AI applications in infrastructure design dates back to the 1980s and refer to the development of expert systems for bridge design (*8*). Neural networks (NNs) and fuzzy set theory have also been implemented in bridge design (*10*, *11*).

TABLE 1 Applications of AI Techniques in Transportation Design

Application Area and Topic	AI Technique	Reference
Highway geometric design:		
Vertical alignment	GA	*31*
	Pattern recognition	*32*
Horizontal alignment	GA	*33*
Vertical–horizontal alignment	GA	*19, 33, 34*
Three-dimensional alignment	GA	*1, 15–18*
	Heuristic search	*35*
Preliminary highway design	GA	*36, 37*
Multiobjective optimization	GA	*38–40*
Pavement design:		
Flexible pavements	Expert system	*41*
Other transportation modes:		
Rail transit station locations	GA	*13, 14*
Railway alignment	GA	*12*
Bridge design	Fuzzy set theory	*11*
	NN	*10, 42*
	Case-based reasoning, Expert system	*43–46*
	NN and GA	*47*
Other:		
Civil engineering elements	Expert system	*48*

Another area of infrastructure design to which several AI techniques have been successfully applied is highway alignment design. GAs and AAs have been used in various applications for optimizing highway alignments (*12*), rail transit alignments (*13*), and for locating stations along rail lines (*14*). Due to the use of GAs and other evolutionary approaches, efficient solutions to the multiobjective optimization of the three-dimensional (3D) highway alignment problem have been developed and tested (*15*, *16*); it is interesting to note that before the application of GAs for 3D alignment optimization, almost all methods applied addressed either vertical or horizontal alignment (*15*). In applying GAs for highway design optimization, instead of using binary digits to represent a solution, a floating point encoding scheme is employed. The chromosome can be easily defined as the set of decision variables, each of which is a continuous real number confined within its associated boundaries (Figure 1). As Kim et al. discuss, GAs may overcome many deficiencies constraining other methods when it comes to geometric alignment applications (*17*). For example, they do not require differentiable objective functions and can model linear and nonlinear relationships and both continuous and discontinuous data space.

An AA application for 3D highway design optimization is described in Jha (*19*). It is an extension of the AA application for the classical traveling salesman problem (TSP) with the main difference being that the ant's travel path are not known a priori, whereas in the TSP city locations through which the ants must pass are known in advance and are fixed. The AAs fall within the domain of swarm intelligence. Swarms are discrete insects with certain characteristics (*20*). Due to the discrete movements of the ants AAs are particularly suited for discrete problems and their application to searching in a continuous space is rather limited.
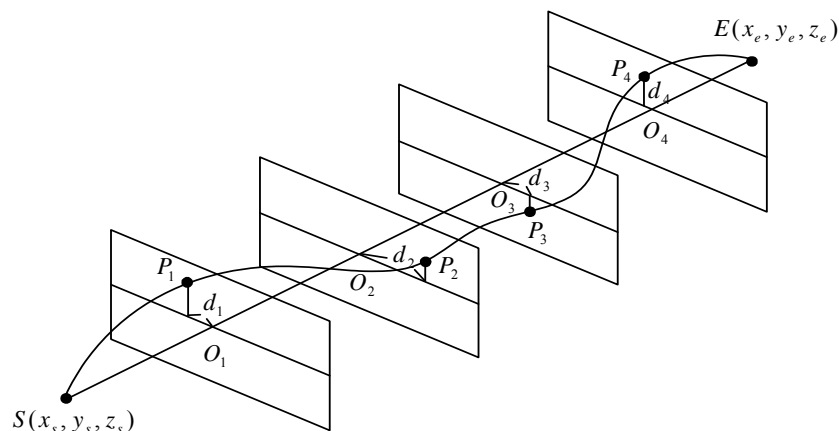
FIGURE 1 Decision variables in the highway design optimization problem (*18*). (Note: S and E are 3-D start and end points, *P*'s are the points of intersection, *O*'s are the points along the straight line connecting *S* and *E*, and d's are the offset distance of *P*'s from *O*'s.)

AI techniques have been used in a variety of applications related to infrastructure construction. The application areas include: (*a*) project evaluation, (*b*) planning and scheduling, (*c*) equipment selection, (*d*) quality control and prediction, (*e*) operation and safety, (*f*) project management, and (*g*) infrastructure maintenance. The specific AI techniques used in these applications are shown in Table 1 and Table 2. The techniques include NN, fuzzy theory, AAs, GAs, expert system, fuzzy and hybrid expert systems (ES), knowledge-based systems, case-based reasoning (CBR), and multiagent systems. The most popular AI techniques are CBR, NN, and ES.

TABLE 2 Applications of AI Techniques in Construction

Application Area and Topic	AI Technique	Reference
Project evaluation:		
Manager selection	Fuzzy GA	*49*
Contract evaluation	Fuzzy ES	*50*
Contractor classification	NN	*24*
Project evaluation	NN	*23*
Planning and scheduling:		
Time–cost trade-off	AA	*51*
	NN	*26*
Construction time and cost	AA	*52*
	CBR	*53–55*
	NN	*25*
	GA	*56*
Scheduling	Knowledge-based system hybrid	*57*
Site layout	ES	*58*
Planning and scheduling	Ant colony	*59*
Path planning	GA	*60*

(*continued*)

TABLE 2 (*continued*) **Applications of AI Techniques in Construction**

Application Area and Topic	AI Technique	Reference
Equipment selection:		
Crane type selection	NN	*27*
Earthmoving equipment selection	GA	*61*
	Knowledge-based system	*62)*
Quality control and prediction:		
Compaction quality estimation	NN	*29*
Project success prediction	Fuzzy theory, NN	*63*
Labor productivity prediction	Fuzzy ES	*64*
Operation and safety:		
Hazard identification	CBR	*22, 65*
Road improvements	CBR	*66*
Earthmoving operations	GA	*67*
	Fuzzy theory	*68*
Bridge inspection	NN–fuzzy sets	*69*
	NN	*70, 71*
	Fuzzy theory	*72*
	Fuzzy theory, NN, GA	*73*
	CBR	*74*
	Neuro-fuzzy	*75, 76*
Project management:		
Dispute resolution	Multiagent systems	*77*
Bid decision making	CBR	*78*
	NN	*28*
Project monitoring	ES	*79*
Selecting contracting strategy	Bayesian networks	*80*
Infrastructure maintenance:		
Multimodal feeder network design	AA	*81*
Categorization of pavement distress	Fuzzy ES	*30*
Inspection and scheduling	GA	*82–84*
Bridge maintenance	CBR	*85*
Roadside appurtenances maintenance	GA	*86*
Pavement maintenance	NN–GA	*87–99*
	CBR	*100*
	GA	*101, 102*
	ES	*103*
	Rough set–NN	*104*

The CBR methodology has originated in AI research to emulate human reasoning and decision-making processes. In construction applications, CBR has been used in many applications such as construction cost estimation and hazard identification and so on (*21*). CBR, which is capable of dealing with qualitative data, has been included in a tool to assist in the worker's daily tasks of identifying hazards and determining appropriate mitigation measures (*22*). The tool prompts users to classify a given work task and then, using a stored library of cases, suggests possible mitigation measures. The user can accept or reject the suggested

strategies. The results are fed back into the system, which becomes more refined for the next work task and the next user.

Moreover, NNs and particularly the Multilayer Perceptron, appear to be rather popular and widely used; they have been applied to project evaluation (*23*) and contractor selection (*24*), project time and cost estimation (*25, 26*), equipment selection (*27*), biding (*28*) and quality estimation (*29*). Moreover, significant research has focused on NN applications in bridge inspection and pavement conditions assessment.

As for ES, the most important task in developing expert systems is to provide a methodology for linking the information gathered from the field and expert knowledge, and the final output generated in a form that would make it amenable to further analysis. Fuzzy theory is a good candidate to tackle the above issues. A typical example is the application proposed by Koduru et al. that used FL and ES techniques for quick, efficient, and consistent classification of flexible pavement distresses using field data. Application of ESs may increase consistency and reduce subjectivity of the classification process (*30*).

Hybrid Systems

Combining AI models for creating hybrid structures is one of the most promising and popular directions for achieving "smarter" and adaptive applications. Hybrid model development can be found in many transportation research fields such as forecasting. The concept is to combine more than one model in order to overcome certain difficulties such as, for example, selecting the appropriate input space, decision making and so on.

In infrastructure design and construction some hybrid models have been employed. Furuta et al. developed a system for damage assessment of reinforced concrete bridge decks that combined GAs and NN to produce fuzzy inference rules (*74*). The weights obtained from the NN are used in the GA valuation function to obtain the optimal combination of rules to be used in the knowledge base of the ES. Abdelrahim and George utilize GAs to optimize the NN performance for selecting the optimum pavement maintenance strategy (*87*). Attoh–Okine used the rough set methodology to reduce the dimension of the pavement management database, before applying artificial NNs, to develop the pavement performance model (*104*).

Srinivas and Ramanjaneyulu, before applying GAs to produce the optimum design solution for a given type of bridge deck, developed a NN to obtain the design responses of T-girder bridge decks which are required for the evaluation of constraints and fitness function in GA (*47*). Ko and Cheng implemented a modeling strategy that considers GAs, NNs, and fuzzy sets. In this hybrid model, a GA searches for the optimum fuzzy rules, optimum NN topology, and optimum parameters of NN (including interconnections status, synaptic weights, bias values, and slopes of activation functions), in a neuro-fuzzy model for dynamic project assessment (*63*). Finally, Li and Burgueño developed a fuzzy ensemble NN (a series of NN models working in parallel) for bridge damage prediction (*69*). In this approach, fuzzy sets are to account for the uncertainty in the modeling and the ensemble structure for the modularity of the problem faced. Modularity typically implies that a problem may be structured into several other sub problems to be modeled more efficiently.

Modeling Comparisons

Although applications of AI in infrastructure design and construction are now quite extensive, studies dedicated to comparing AI model performance to other modeling approaches are limited. Most studies attempt to establish a certain benchmark study by comparing different AI techniques (fuzzy sets FS with NNs as an example), for a specific problem set. A typical example can be found in the hybrid models developed by Ko and Cheng and Li and Burgueño, where the authors compare the hybrid models they developed with the performance of the same AI techniques (*63*, *69*). Ko and Cheng found that the genetically optimized fuzzy NN methodology outperformed pure FL and pure NNs as the proposed model automatically identifies all parameters required in a fuzzy NN (*63*). The same applied to the model proposed by Li and Burgueño (*69*).

Comparing the performance of AI models to classical statistical models is, surprisingly, not popular. A good example is the work of Yang et al. which compared the recurrent Markov chain and NN models for multiple-year prediction of crack performance. The authors found that the recurrent Markov chain tends to produce more consistent forecasts as compared to NNs which exhibit a tendency to overpredict crack deterioration. Moreover, statistical inference is hardly ever used to test the efficiency of the models developed (*6*, *98*). An exception is the work of Jong et al. and Jong and Schonfeld that developed an experiment to statistically test the goodness of the solution of their GA in optimizing 3D highway alignments (*15*, *36*).

Most interdisciplinary applications of AI have for long emphasized the need for synergies between AI and statistical inference for increasing modeling efficiency (*6*). This was proposed for two reasons. First, AI applications can be thought of as typical experiments that should be evaluated against optimum design, modeling structure and performance of results; indeed, classical statistics may provide a well founded and much needed mathematical platform for such evaluations. Second, causation—a desired attribute in engineering applications—is not straightforwardly tackled by AI techniques, but can be extracted using statistical modeling. Although AI models can easily model complex datasets with missing values, latent variables, and temporal effects that cannot be well approximated by simple nonrecursive statistical models, statistics can effectively help to assess the characteristics of the input–output relationships and shed light to the AI models.

DISCUSSION AND FUTURE OUTLOOK

From the preceding discussion it is clear that AI techniques have found widespread application in the area of infrastructure design and construction. It is also clear that many of these approaches have been used in a relative vacuum, in that they are most frequently used to test a particular phenomenon, however not within an integrated system. It appears that new AI approaches are rarely, if ever, compared to other more established and widely used techniques. This is a pronounced shortcoming since in the area of infrastructure design and construction the risks of failure are very costly, while the approaches currently used have been thoroughly tested for many decades. As such, the process of establishing the superiority of a new method over established ones invariably goes through the comparison route. From the literature in this area it is important to note that, in practice, AI systems should be developed as decision-support tools rather than assuming that AI systems could be developed to replace human expertise. It is also

very important to note that there is an increasing tendency for developing hybrid systems that mix different AI techniques. This is a promising avenue for future research as the advantages of different approaches can be combined, while the disadvantages minimized.

We note that many AI techniques can be straightforwardly applied for solving a number of construction-related problems. However, many other approaches are still widespread because of their proven results and years of use. The scientific literature in the area of planning and construction has, over the past few years, shown an obvious trend in the mathematical tools used ranging from simpler, linear, low-dimensional to complex, nonlinear, high-dimensional systems, largely because of significant increases in computing power, but still not necessarily justified by logic or fundamental research needs. We believe that the problem to be solved is more important than the tools used and that complex, nonlinear tools have both advantages and limitations. That is, frequently, simpler models give results that are easier to understand and use in practice than the results obtained from complex models.

REFERENCES

1. Kang, M. W., P. Schonfeld and N. Yang. Prescreening and Repairing in a Genetic Algorithm for Highway Alignment Optimization. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 24, 2009, pp. 109–119.
2. Funderburg, R. G., H. Nixon, M. G. Boarnet, and G. Ferguson. New Highways and Land Use Change: Results from a Quasi-Experimental Research Design. *Transportation Research Part A*, Vol. 44, 2010, pp. 76–98.
3. Sundin, S., and C. Braban-Ledoux. Artificial Intelligence−Based Decision Support Technologies in Pavement Management. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 16, 2001, pp. 143–157.
4. Negenborn, R. R., and H. Hellendoorn. Intelligence in Transportation Infrastructures via Model-Based Predictive Control, Intelligent Infrastructures. *Intelligent Infrastructures, Intelligent Systems, Control and Automation: Science and Engineering* (R. R. Negenborn, Z. Lukszo, and H. Hellendoorn, eds.), Vol. 42, Part 1, Springer Netherlands, 2010, pp. 3–24.
5. Flintsch, G. W., and C. Chen. Soft Computing Applications in Infrastructure Management. *Journal of Infrastructure Systems*, Vol. 10, No. 4, 2004, pp. 157–166.
6. Karlaftis, M. G., and E. I. Vlahogianni. Statistics versus Neural Networks in Transportation Research: Differences, Similarities and Some Insights. *Transportation Research Part C*, Vol. 19, 2011, pp. 387–399.
7. Sadek, A. W. Artificial Intelligence Applications in Transportation. In *Transportation Research Circular E-C113: Artificial Intelligence in Transportation: Information for Application*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 1–6.
8. Reich, Y. Machine Learning Techniques for Civil Engineering Problems. *Microcomputers in Civil Engineering,* Vol. 12, 1997, pp. 295–310.
9. Adeli, H. Neural Networks in Civil Engineering: 1989–2000. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 16, 2001, pp. 126–142.
10. Zhao, Z., W. He, and S. C. Fan. Preliminary Design System for Concrete Box Girder Bridges. *Journal of Computing in Civil Engineering*, Vol. 15, No. 3, 2001, pp. 184–192.
11. Malekly, H., S. M. Mousavi, and H. Hashemi. A Fuzzy Integrated Methodology for Evaluating Conceptual Bridge Design. *Expert Systems with Applications*, Vol. 37, 2010, pp. 4910–4920.
12. Jha, M. K., P. Schonfeld, and S. Samanta. Optimizing Rail Transit Routes with Genetic Algorithms and Geographic Information System. *Journal of Urban Planning and Development*, Vol. 133, No. 3, 2007, pp.161–171.

13. Samanta, S., and M. K. Jha. Modeling a Rail Transit Alignment Considering Different Objectives. *Transportation Research Part A*, Vol. 45, No. 1, 2011, pp. 31–45.

14. Samanta, S., and M. K. Jha. Identifying Feasible Locations for Rail Transit Stations: Two-Stage Analytical Model. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2063*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 81–88.

15. Jong, J.-C., and P. Schonfeld. An Evolutionary Model for Simultaneously Optimizing 3-Dimensional Highway Alignments. *Transportation Research*, Vol. 37, No. 2, 2003, pp. 107–128.

16. Jha, M. K., and P. Schonfeld. Integrating Genetic Algorithms and Geographic Information System to Optimize Highway Alignments. *Transportation Research Record: Journal of the Transportation Research Board, No. 1719*, TRB, National Research Council, Washington, D.C., 2000, pp. 233–240.

17. Kim, E., M. K. Jha, and B. Son. Improving the Computational Efficiency of Highway Alignment Optimization Models Through a Stepwise Genetic Algorithms Approach. *Transportation Research*, Vol. 39, No. 4, 2005, pp. 339–360.

18. Jha, M. K., P. Schonfeld, J.-C. Jong, and E. Kim. *Intelligent Road Design*. WIT Press, Boston, Mass., 2006.

19. Jha, M. K. A Geographic Information Systems-Based Model for Highway Design Optimization. Ph.D. dissertation. University of Maryland, College Park, 2000.

20. Teodorovic, D. Swarm Intelligence Systems for Transportation Engineering: Principles and Applications. *Transportation Research Part C*, Vol. 16, No. 6, 2008, pp. 651–667.

21. Sadek, A. W. Hybrid Simulated Annealing and Case-Based Reasoning Approach for Computationally Intensive Transportation Problems: Rationale and Design Issues. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1774*, TRB, National Research Council, Washington, D.C., 2001, pp. 18–24.

22. Campbell, J. M., S. D. Smith, M. C. Forde, and R. D. Ladd. Identifying Hazards in Transportation Construction and Maintenance Tasks: Case-Based Reasoning Approach Using Railroad Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1995*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 69–75.

23. Lu, M. Enhancing Project Evaluation and Review Technique Simulation through Artificial Neural Network-Based Input Modeling. *Journal of Construction Engineering and Management*, Vol. 128, 2002, pp. 438–445.

24. Elazouni, A. M. Classifying Construction Contractors Using Unsupervised-Learning Neural Networks. *Journal of Construction Engineering and Management*, Vol. 132, 2006, pp. 1242–1253.

25. Hegazy, T., and A. Ayed. Neural Network Model for Parametric Cost Estimation of Highway Projects. *Journal of Construction Engineering and Management*, Vol. 124, No. 3, 1998, pp. 210–219.

26. Sodikov, J. Road Cost Models for Prefeasibility Studies in Developing Countries. *Journal of Infrastructure Systems*, Vol. 15, No. 4, 2009, pp. 278–289.

27. Sawhney, A., and A. Mund. Adaptive Probabilistic Neural Network-Based Crane Type Selection System. *Journal of Construction Engineering and Management,* Vol. 128, 2002, pp. 265–273.

28. Williams, T. P. Bidding Ratios to Predict Highway Project Costs. *Engineering, Construction and Architectural Management*, Vol. 12, No. 1, 2005, pp. 38–51.

29. Commuri, S., A. T. Mai, and M. Zaman. Neural Network-Based Intelligent Compaction Analyzer for Estimating Compaction Quality of Hot Asphalt Mixes. *Journal of Construction Engineering and Management*, Vol. 137, 2011, pp. 634–644.

30. Koduru, H. K., F. Xiao, S. N. Amirkhanian, and C. H. Juang. Using Fuzzy Logic and Expert System Approaches in Evaluating Flexible Pavement Distress: Case Study. *Journal of Transportation Engineering*, Vol. 136, No. 4, 2010, pp. 149–157.

31. Fwa, T. W., W. T. Chan, and Y. P. Sim. Optimal Vertical Alignment Analysis for Highway Design. *Journal of Transportation Engineering*, Vol. 128, No. 5, 2002, pp. 395–402.

32. Easa, S., and F. Wang. Estimating Continuous Highway Vertical Alignment Using the Least Squares Method. *Canadian Journal of Civil Engineering*, Vol. 37, 2010, pp. 1362–1370.

33. Jong, J.-C. Optimizing Highway Alignments with Genetic Algorithms. Ph.D. dissertation. University of Maryland, College Park, 1998.
34. Kim, E. Modeling Intersections and Other Structures in Highway Alignment Optimization. Ph.D. dissertation. University of Maryland, College Park, 2001.
35. Mandow, L., and J. L. de la Perez. SINDI: An Intelligent Assistant for Highway Design. *Expert Systems with Applications*, Vol. 27, No. 4, 2004, pp. 635–644.
36. Jong, J.-C., M. K. Jha, and P. Schonfeld. Preliminary Highway Design with Genetic Algorithms and Geographic Information Systems. *Computer-Aided Civil Infrastructure Engineering*, Vol. 15, No. 4, 2000, pp. 261–271.
37. Jha, M. K., C. McCall, and P. Schonfeld. Using GIS, Genetic Algorithms, and Computer Visualization in Highway Development. *Computer-Aided Civil Infrastructure Engineering,* Vol. 16, No. 6, 2001, pp. 399–414.
38. Maji, A., and M. K. Jha. Multi-Objective Highway Alignment Optimization Using a Genetic Algorithm. *Journal of Advanced Transportation*, Vol. 43, 2009, pp. 481–504.
39. Kim, E., M. K. Jha, D. J. Lovell, and P. Schonfeld. Intersection Modeling for Highway Alignment Optimization. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 19, 2004, pp.119–129.
40. Kim, E., M. K. Jha, P. Schonfeld, and H. S. Kim. Highway Alignment Optimization Incorporating, Bridges and Tunnels. *Journal of Transportation Engineering*, Vol. 133, No. 2, 2007, pp. 71–81.
41. Goh, A. T. Advisory Expert System for Flexible Pavement Design. *Artificial Intelligence in Engineering*, Vol. 8, 1993, pp. 47–56.
42. Hong, N. K., S.-P. Chang, and S.-C. Lee. Development of ANN-Based Preliminary Structural Design Systems for Cable-Stayed Bridges. *Advances in Engineering Software*, Vol. 33, No. 2, 2002, pp. 85–96.
43. Welch, J. G., and M. Biswas. Application of Expert Systems in the Design of Bridges. In *Transportation Research Record 1072*, TRB, National Research Council, Washington, D.C., 1986, pp. 65–70.
44. Biswas, M., and J. G. Welch. BDES: A Bridge Design Expert System. *Engineering with Computers,* Vol. 2, No. 3, 1987, pp. 125–136.
45. Burgoyne, C. J., and S. H. R. Sham. Application of Expert Systems to Prestressed Concrete Bridge Design. *Civil Engineering and Environmental Systems*, Vol. 4, No. 1, 1987, pp. 14–19.
46. Choi, C. K., and I. H. Choi. An Expert System for Selecting Types of Bridges. *Computers and Structures*, Vol. 48, No. 2, 1993, pp. 183–192.
47. Srinivas, V., and K. Ramanjaneyulu. An Integrated Approach for Optimum Design of Bridge Decks Using Genetic Algorithms and Artificial Neural Networks. *Advances in Engineering Software*, Vol. 38, No. 7, 2007, pp. 475–487.
48. Pérez de la Cruz, J. L., R. Conejo-Muñoz, R. Morales-Bueno, and J. Puy-Huarte. Highway Design by Constraint Specification. *Artificial Intelligence in Engineering*, Vol. 9, No. 2, 1995, pp. 127–139.
49. Rashidi, A., F. Jazebi, and I. Brilakis. Neurofuzzy Genetic System for Selection of Construction Project Managers. *Journal of Construction Engineering and Management*, Vol. 137, No. 1, 2011, pp. 17–29.
50. Marsh, K., and A. R. Fayek. SuretyAssist: Fuzzy Expert System to Assist Surety Underwriters in Evaluating Construction Contractors for Bonding. *Journal of Construction Engineering and Management*, Vol. 136 , No. 11, 2010, pp. 1219–1226.
51. Afshar, A., A. K. Ziaraty, A. Kaveh, and F. Sharifi. Nondominated Archiving Multicolony Ant Algorithm in Time–Cost Trade-Off Optimization. *Journal of Construction Engineering and Management*, Vol. 135, No. 7, 2009, pp. 668–674.
52. Ng, S. T., and Y. Zhang. Optimizing Construction Time and Cost Using Ant Colony Optimization Approach. *Journal of Construction Engineering and Management*, Vol. 134, No. 9, 2008, pp. 721–728.
53. Kim, G.-H., S.-H. An, and K.-I. Kang. Comparison of Construction Cost Estimating Models Based on Regression Analysis, Neural Networks, and Case-Based Reasoning. *Building and Environment*, Vol. 39, No. 10, 2004, pp. 1235–1242.
54. Al-Tabtabai, H., A. P. Alex, and M. Tantash. Preliminary Cost Estimation of Highway Construction Using Neural Networks. *Cost Engineering*, Vol. 41, No. 3, 1999, pp. 19–24.

55. Jaśkowski, P., and A. Sobotka. Scheduling Construction Projects Using Evolutionary Algorithm. *Journal of Construction Engineering and Management*, Vol. 118, 2006, pp. 861–870.

56. Hsieh, T.-Y., and H.-L. Liu. Genetic Algorithm for Optimization of Infrastructure Investment Under Time-Resource Constraints. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 19, No. 3, 2004, pp. 203–12.

57. Tommelein, I. D., R. E. Levitt, and B. Hayes-Roth. SightPlan Model for Site Layout. *Journal of Construction Engineering and Management*, Vol. 118, 1992, pp. 749–766.

58. Moselhi, O., and M. J. Nicholas. Hybrid Expert System for Construction Planning and Scheduling. *Journal of Construction Engineering and Management*, Vol. 116, 1990, pp. 221–238.

59. Vitins, B. J., and K. W. Axhausen. Optimization of Large Transport Networks Using the Ant Colony. *Heuristic Computer-Aided Civil and Infrastructure Engineering*, Vol. 24, 2008, pp. 1–14.

60. Soltani, A. R., H. Tawfik, J. Y. Goulermas, and T. Fernando. Path Planning in Construction Sites: Performance Evaluation of the Dijkstra, Ap, and GA Search Algorithms. *Advanced Engineering Informatics*, Vol. 16, 2002, pp. 291–303.

61. Haidar, A., S. Naoum, R. Howes, and J. Tah. Genetic Algorithms Application and Testing for Equipment Selection. *Journal of Construction Engineering and Management*, Vol. 125, 1999, pp. 32–38.

62. Amirkhanian, S. N., and M. J. Baker. Expert System for Equipment Selection for Earth-Moving Operations. *Journal of Construction Engineering and Management*, Vol. 118, No. 2, 1992, pp. 318–331.

63. Ko, C.-H., and M.-Y. Cheng. Dynamic Prediction of Project Success Using Artificial Intelligence. *Journal of Construction Engineering and Management*, Vol. 133, 2007, pp. 316–322.

64. Fayek, A. R., and A. Oduba. Predicting Industrial Construction Labor Productivity Using Fuzzy Expert Systems. *Journal of Construction Engineering and Management*, Vol. 131, 2005, pp. 938–941.

65. Goh, Y. M., and D. K. Chua. Case-Based Reasoning Approach to Construction Safety Hazard Identification: Adaptation and Utilization. *Journal of Construction Engineering and Management,* Vol. 136, 2010, pp. 170–178.

66. Lin, F., T. Sayed, and P. Deleur. Estimating Safety Benefits of Road Improvements: Case Based Approach. *Journal of Transportation Engineering*, Vol. 129, No. 4, 2003, pp. 385–391.

67. Moselhi, O., and A. A. Adel. Optimization of Earthmoving Operations in Heavy Civil Engineering Projects. *Journal of Construction Engineering and Management*, Vol. 135, No. 10, 2009, pp. 948–954.

68. Göktepe, A. B., A. H. Lav, S. Altun, and G. Altintaş. Fuzzy Decision Support System to Determine Swell/Shrink Factor Affecting Earthwork Optimization of Highways. *Mathematical and Computational Applications*, Vol. 13, No. 1, 2008, pp. 61–70.

69. Li, Z., and R. Burgueño. Using Soft Computing to Analyze Inspection Results for Bridge Evaluation and Management. *Journal of Bridge Engineering*, Vol. 15, 2010, pp. 430–438.

70. Cattan, J., and J. Mohammadi. Analysis of Bridge Condition Rating Data Using Neural Networks. *Microcomputing in Civil Engineering,* Vol. 12, No. 6, 1997, pp. 419–429.

71. Sobanjo, J. *Proc., 4th Congress on Computing in Civil Engineering*, ASCE, New York, 1997, pp. 623–626.

72. Molina, A., and K. Chou. Evaluation of Existing Bridges Using Neural Networks. *Structural Engineering and Mechanics*, Vol. 13, No. 2, 2002, pp. 187–209.

73. Tee, A., M. Bowman, and K. Sinha. Application of Fuzzy Logic to Condition Assessment of Concrete Slab Bridges. In *Transportation Research Record, 1184*, TRB, National Research Council, Washington, D.C., 1988, pp. 22–30.

74. Furuta, H., J. He, and E. Watanabe. A Fuzzy Expert System for Damage Assessment Using Genetic Algorithms and Neural Networks. *Microcomputers in Civil Engineering*, Vol. 11, No. 1, 1996, pp. 37–45.

75. Roddis, W. M. K., and J. Bocox. Case-Based Approach for Steel Bridge Fabrication Errors. *Journal of Computing in Civil Engineering*, Vol. 11, No. 2, 1997, pp. 84–91.

76. Wang, Y.-M., and T. M. S. Elhag. An Adaptive Neuro-Fuzzy Inference System for Bridge Risk Assessment. *Expert Systems with Applications*, Vol. 34, No. 4, 2008, pp. 3099–3106.

77. El-adaway, I., and A. A. Kandil. Multiagent System for Construction Dispute Resolution (MAS-COR). *Journal of Construction Engineering and Management*, Vol. 136, 2010, pp. 303–315.

78. Chua, D. K., D. Z. Li, and W. T. Chan. Case-Based Reasoning Approach in Bid Decision Making. *Journal of Construction Engineering and Management*, Vol. 127, 2001, pp. 35–46

79. McGartland, M. R., and C. T. Hendrickson. Expert Systems for Construction Project Monitoring. *Journal of Construction Engineering and Management*, Vol. 111, 1985, pp. 293–307.

80. Bayraktar, M. E., and M. Hastak. A Decision Support System for Selecting the Optimal Contracting Strategy in Highway Work Zone Projects. *Automation in Construction*, Vol. 18, No. 6, 2009, pp. 834–843.

81. Mohaymany, A. S., and A. Gholami. Multimodal Feeder Network Design Problem: Ant Colony Optimization Approach. *Journal of Transportation Engineering*, Vol. 136, No. 4, 2010, pp. 323–331.

82. Maji, A., and M. K. Jha. Modeling Highway Infrastructure Maintenance Schedule with Budget Constraint. *Transportation Research Record: Journal of the Transportation Research Board, No. 1991*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 19–26.

83. Jha, M. K. Optimal Highway Infrastructure Maintenance Scheduling Considering Deterministic and Stochastic Aspects of Deterioration. *Sustainable and Resilient Critical Infrastructure Systems* (K. Gopalakrishnan and S. Peeta, eds.), Springer, 2010, pp. 231–248.

84. Jha, M. K., K. Kepaptsoglou, M. Karlaftis, and J. Abdullah. A Genetic Algorithms-Based Decision Support System for Transportation Infrastructure Management in Urban Areas. *Recent Advances in City Logistics: Proceedings of the 4th International Conference on City Logistics* (E. Taniguchi and R. Thompson, eds.), Elsevier Publishing Company, 2006, pp. 509–523.

85. Morcous, G., H. Ricard, and A. M. Hanna. Modeling Bridge Deterioration Using Case Based Reasoning. *Journal of Infrastructure Systems*, Vol. 8, No. 3, 2002, pp. 86–94.

86. Jha, M. K., and J. Abdullah. A Markovian Approach for Optimizing Highway Life-Cycle with Genetic Algorithms by Considering Maintenance of Roadside Appurtenances. *Journal of the Franklin Institute*, Vol. 343, 2006, pp. 404–419.

87. Abdelrahim, A. M., and K. P. George. Artificial Neural Network for Enhancing Selection of Pavement Maintenance Strategy. In *Transport Research Record: Journal of the Transportation Research Board, No. 1669*, TRB, National Research Council, Washington, D.C., 2000, pp. 16–22.

88. Huang, Y., and R. K. Moore. Roughness Level Probability Prediction Using Artificial Neural Networks. In *Transportation Research Record 1592*, TRB, National Research Council, Washington, D.C., 1997, pp. 89–97.

89. Fwa, T. F., W. T. Chan, and C. T. Lim. Decision Framework For Pavement Friction Management Of Airport Runways. *Journal of Transportation Engineering*, Vol. 123–126, 1997, pp. 429–435.

90. Owusu-Ababia, S. Effect of Neural Network Topology on Flexible Pavement Cracking Prediction. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 13, No. 5, 1998, pp. 349–355.

91. Shekharan, A. R. Effect of Noisy Data on Pavement Performance Prediction by Artificial Neural Networks. In *Transportation Research Record 1643*, TRB, National Research Council, Washington, D.C., 1998, pp. 7–13.

92. Roberts, C. A., and N. O. Attoh-Okine. A Comparative Analysis of Two Artificial Neural Networks Using Pavement Performance Prediction. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 13, 1998, pp. 339–348.

93. Lou, Z., M. Gunaratne, J. J. Lu, and B. Dietrich. Application of a Neural Network Model to Forecast Short-Term Pavement Crack Condition: Florida Case Study. *Journal of Infrastructure Systems*, Vol. 7, No. 4, 2001, pp. 166–174.

94. Mukkamala, S., and A. H. Sung. Feature Selection for Intrusion Detection using Neural Networks and Support Vector Machines. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1823*, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 33–39.

95. Bosurgi, G., A. D'Andrea, and F. Trifirò. Development of a Sideway Force Coefficient Prediction Model Based on the Artificial Neural Networks. Presented at 2nd European Pavement and Asset Management Conference, Berlin, Germany, 2004.

96. Lee, B. J., and H. Lee. Position Invariant Neural Network for Digital Pavement Crack Analysis. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 12, No. 2, 2004, pp. 105–108.

97. Bosurgi, G., and F. Trifirò. A Model Based on Artificial Neural Networks and Genetic Algorithms for Pavement Maintenance Management. *International Journal of Pavement Engineering*, Vol. 6, No. 3, 2005, pp. 201–209.

98. Loizos, A., and M. Karlaftis. Neural Networks and Non-Parametric Statistical Models: A Comparative Analysis in Pavement Condition Assessment. *Journal Advances and Applications in Statistics*, Vol. 6, No. 1, 2006, pp. 87–110.

99. Yang, J., J. J. Lu, M. Gunaratne, and B. Dietrich. Modeling Crack Deterioration of Flexible Pavements: Comparison of Recurrent Markov Chains and Artificial Neural Networks. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1974*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 18–25.

100. Chou, J. S. Web-Based CBR System Applied to Early Cost Budgeting for Pavement Maintenance Project. *Expert Systems with Applications*, Vol. 36, 2009, pp. 2947–2960.

101. Unnikrishnan, A., V. Valsaraj, I. Damnjanovic, and T. S. Waller. Design and Management Strategies for Mixed Public Private Transportation Networks: A Meta-Heuristic Approach. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 24, No. 4, 2009, pp. 266–279.

102. Chew, R. L., Y. Wang, and T. F. Fwa. Genetic Algorithm-Simulation Methodology For Pavement Maintenance Scheduling. *Computer-Aided Civil Infrastructure Engineering*, Vol. 19, 2004, pp. 446–455.

103. Ritchie, S. G. Expert Systems in Pavement Management. *Transportation Research Part A*, Vol. 21A , No. 2, 1987, pp. 145–152.

104. Attoh-Okine, N. O. Combining Use of Rough Set and Artificial Neural Networks in Doweled-Pavement-Performance Modeling-A Hybrid Approach. *Journal of Transportation Engineering*, Vol. 128, No. 3, 2002, pp. 270–275.

Thoughts on the Future of
Artificial Intelligence and Transportation

Thoughts on the Future of
Artificial Intelligence and Transportation

I n this concluding part of the circular, members of the committee share their personal thoughts on the future of artificial intelligence (AI) and transportation. The brief remarks are organized alphabetically by author.

MONTASIR ABBAS
Virginia Tech

Application of AI in transportation is receiving a growing attention in the new era of rich and detailed data. One of the main areas that I expect to flourish in the near future is agent-based modeling and simulation (ABMS). The transportation profession is moving towards collecting richer sets of data that open the door for more in-depth analysis of phenomena that we only dreamed to conceptualize in the past. One of the major challenges that we will soon face is the development of new ABMS techniques to explore new concepts. Traditional AI techniques (e.g., artificial neural networks, genetic algorithms, fuzzy logic) are expected to lead the way, at least initially, in the ABMS exploration arena. Other techniques will emerge and researchers will be burdened to prove their worth (the techniques, not the researchers) before they become mainstream methods.

EREL AVINERI
University of the West of England

Much of the recent work in the field of travel behavior focuses on the cognitive biases and the bounded rationality of travelers exhibited in route, mode, and travel time choices. Some deviations from rational models of travel choice were found to be systematic, consistent, robust, and largely predictable. One of the aims of AI applications is to demonstrate mechanisms that underlie human behaviour. The cognitive biases, flaws, and limitations of "real" humans exhibited in the processes of judging, inferencing, learning, and problem solving could be incorporated in the development of the next generation of AI methodologies and tools.

RYAN FRIES
Southern Illinois University, Edwardsville

The challenges faced by today's transportation community are ever increasing in their complexity. To address these issues, long-range research must address key diverging interests such as sustainability and roadway capacity. The depth and breadth of such work requires advanced tools such as AI to identify sets of optimal solutions. The primary challenge lies with representing the intricacies of long-term planning goals and the realities of the decision-making process with accuracy.

Further, continued trends such as globalism influence transportation on a variety of fronts. For example, the cycle of identifying new threats and creating countermeasures significantly influences the decision-making process and the subsequent funding. Although there will always be some risk of security incidents, accurate information guiding the selection of mitigation projects can help decision makers select from projects that promise the highest returns.

Unfortunately, identifying optimal solutions to the profound sustainability or security challenges facing the transportation industry presents a complex problem. There has never been a better opportunity to use AI and advanced computing to identify feasible paths toward mitigating today's transportation challenges.

SHERIF ISHAK
Louisiana State University

The field of AI continues to expand and support a wide spectrum of applications in all areas of transportation engineering. The open literature reveals a surge of new applications that successfully invoked modeling approaches from the area of AI in the past two decades. Such applications show that the strength of AI techniques stems from the inherently advanced computing algorithms that are highly required to address the complexity of high-dimensional research problems, as well as the intrinsic uncertainties associated with such problems. While research funding in the area of AI was cut back in recent years, the general perception of the research community appears to have shifted from viewing AI techniques as a black box to a better understanding of their strengths and limitations. This is evident with the emergence of new AI paradigms in the last decade. Historically, AI was classified into two approaches: a classical one involving design of AI and based on symbolic reasoning and a connectionist approach involving development of AI based on artificial neural networks and genetic algorithms. But in the last decade, more AI-based techniques and algorithms were developed to cope with challenges that conventional AI methods could not overcome. Today, more researchers and practicing engineers continue to raise their expectations in finding solutions to theoretical and practical problems in AI paradigms. Examples of such applications are found in the area of information extraction and data mining, development of performance measures, advanced traffic management functions (incident detection, traffic prediction, etc.), and optimization and calibration of large scale networks, to name a few. As engineering inspired by science, the field of AI is far from fulfilling its promise or reaching the end of its spectrum. It needs to grow and re-group continually to rise to its potential in solving complex problems that we face today in applications of transportation engineering and other fields of research.

MANOJ JHA
Morgan State University

I see AI as a great tool for solving many interesting and complex transportation problems, such as vehicle routing problems, highway design and route optimization, and infrastructure maintenance and management. Having reviewed and written many papers dealing with AI applications in transportation I have come to the conclusion that most of the contributors do not

necessarily understand the theoretical foundations of the AI algorithms being employed to seek the optimal solution. The knowledge of the theoretical foundations, including the coefficients, parameters, and other factors that influence the optimal search process, computation efficiency, and quality of solution is important. Therefore, these algorithms should not be treated as a plug-and-play device, rather a well thought out synergy between the problem and solution algorithm is extremely important.

As an example, I would like to talk about the foundations of genetic algorithms (GAs). The schema theorem and the building block hypothesis attempt to explain the power of GAs in terms of how schemata are processed. The schema theorem states that schemata with short defining length, low order, and better fitness (called building blocks) receive exponentially increasing trials in subsequent generations of a GA. The building block hypothesis says that a GA seeks near-optimal performance through the juxtaposition of building blocks by genetic operators. However, these two issues have been critically argued and challenged in the last decade. Opponents of schema theory argue that it tells us very little about what is really going on inside the optimal search. Moreover, the traditional Holland–Goldberg schema theorem is pessimistic in the sense that it provides only a lower bound on expected schema growth. Further, it was developed for GAs with fixed-length, binary representation using standard GA genetic operators. Therefore, caution should be exercised in employing AI algorithms since otherwise both computation efficiency and quality of solution may be compromised. As a final note, not all algorithms are suitable for searching in a continuous search space. For example, ant colony optimization does a great job searching in a discrete space, but they are not suited for continuous spaces. On the other hand, GAs are suitable for searching in both discrete and continuous spaces."

SHINYA KIKUCHI
Virginia Tech

Our world is complex; many decision parameters are involved, and they are related in a very complicated manner. A change in the state of one parameter affects many parts of the system, and the impacts are often unpredictable. The traditional paradigms of prediction, diagnosis, and regulation or optimization are not sufficient to deal with the extremely complex social and human systems, in which transportation engineering and planning is part of.

AI offers various tools to deal with problems related to optimization, classification, control, search, inference, knowledge acquisition, and data treatment. What is most interesting to me is that many of these AI tools can be "mixed" in parallel and serial connections as the analyst sees them fit to the problem context. In other words, AI allows imagination and creation in the use of tools. These characteristics of AI free our mind and attitude from the traditional mindset that models must be based on the rigid air-tight mathematical construct.

The potential risk of AI use, however, is that the tools are often loosely constructed; some are intuitive and mathematical integrity is not completely proven. Hence, outcome of the analysis must be checked and interpreted carefully for their accuracy and the acceptable bounds of the solution. Yet, this checking process provides us with the opportunity to think philosophically about the nature of the transportation problems, the nature of the analytical process, and the nature of the solution being sought and its acceptability. After all, solutions to a transportation

problem are many. Hence, an approach that will produce different and creative solutions is what we should look for.

HONGCHAO LIU
Texas Tech University

Since 1956 when McCarty first proposed the term of AI, the advancement of AI in the past 60 years has witnessed a series of creative methods, algorithms, and theories, which have been successfully adopted in many engineering fields to address real-world problems.

Although usually regarded as a research branch under computer science, AI has expanded its research focus from the brains and cognitive reasoning to a wide range of organisms, processes, and phenomena at multiple spatial and temporal scales. This diversity across multiple research areas revives AI from the "AI Winter" in 1970s to the success in the new millennium.

The primary motivation for the future of AI will still lay in the inspirations and integrations from relative research areas especially the frontiers of physiology and biology. The biological observations and studies inspire the researchers to establish systems that mimic the behaviors of animals and human. On the spatial scale, bio-inspired AI ranges from swarm intelligence, ant colony optimization to cellular automata. On the temporal scale, it spans from evolutionary computation, neural networks to artificial immune systems. From the trend and challenges of recent developments in bio-inspired AI, following three very promising directions can be foreseen:

- *Bio-inspired computation on a large-scale distributed architecture*. Bio-inspired computation algorithms, especially the collective intelligence, are derived from the success of collective behavior, where the individuals follow a certain pattern of interactions. Due to this advantage, bio-inspired AI is very suitable for the cooperative optimization on a distributed architecture. Recent researches are shifting more and more of its focus onto the cooperation and negotiation mechanism for distributed bio-inspired AI.
- *Fusion of collective and evolutionary intelligence*. Collective intelligences, like ant colony optimization, are inspired from the spatial activities of animals, while the evolutionary algorithm and the immune system are inspired from the temporal perspective. If introducing the adaptation of evolution into the interactive rules, the collective intelligence may tend to be stable and dependable over iterations. This spatial–temporal fusion between the collective and evolutionary intelligences has great potential and received increasing attention recently.
- *Theoretical foundation for bio-inspired computations*. One of the challenges that hinder further development of bio-inspired AI is lack of theoretical foundation. Due to the loosely coupled processes, the performance of bio-inspired is not guaranteed in a defined domain. However, in recent years, researchers are making more efforts to establish a solid theoretical foundation for bio-inspired computation and the theory including run-time analysis, population dynamics, and fitness landscape analysis has grown rapidly.

These potential breakthroughs in AI will provide professionals in transportation engineering with more effective and efficient methods for travel behavior modeling under chaotic conditions such as evacuation. Applications may include, but are not limited to car-following behavior analysis, route choice modeling, and traffic assignment in a traffic network

with uncertainty. It could also lead to real-time control algorithms with greater computational efficiency, which is critical to real-time traffic control for large networks. For example, bio-inspired AI on a distributed computing architecture will transform current central-controlled operations to a loosely coupled distributed management. This is very promising especially for the coordinated control of traffic signals in large-size traffic networks, where the traditional optimization methods fail to provide a feasible and efficient solution. The bio-inspired AI method regards the traffic system optimization problem as an interactive evolution in a collective society. The distributed operators, as the individuals in collective intelligence, will not only focus on their own optima, but also seek cooperation to achieve the system optima. Meanwhile, the adaptation from evolutionary computations would also benefit to the real-time adjustment of the operation strategy.

EDARA PRAVEEN
University of Missouri

The fields of AI and advanced computing provide immense value to transportation research. Their applications in mining traffic datasets, forecasting traffic parameters, explaining traffic phenomena, and intelligent real-time traffic control, are a few examples of their valuable contributions that I have personally explored. Like statistical and econometric modeling, the algorithms and methods in the area of artificial intelligence and advanced computing are inherently complex and often involve rigorous mathematical underpinnings. Over the years, researchers have applied different AI techniques and algorithms to transportation areas and thereby making these techniques accessible to a wide range of audience in transportation. They have developed several decision-support tools, written articles and books to enhance the use of these techniques in transportation. The TRB committee on AI and advanced computing strives to continue making contributions to transportation research through its annual workshops, conference presentations and publications, and outreach activities.

The future of AI applications to transportation holds a lot of promise given the advances in computing such as Cloud computing and Web 2.0 combined with the needs for increased traffic data archiving and real-time information dissemination. Building the next generation of data-driven web-based traffic simulation models is another area where AI and advanced computing techniques could play a key role in the near future.

YI QI
Texas Southern University

In the future, traffic management system will be more and more computer based to meet the requirements of speed, comfort, and safety. Thus, AI techniques will be widely applied in real-time traffic management to ease freeway congestion and improve roadway safety. AI techniques, such as neural work, fuzzy logic, and knowledge-based expert system, will be applied to provide traffic management center decision-support tools and advanced surveillance and communication functions. The typical applications will include advanced incident detection and response system, hazard evaluation system, and real-time traveler information system.

In addition, with the emergence of vehicle infrastructure integration concept, AI techniques will have the potential applications in "smart" cars and automated highways. A next generation of transport will be without drivers. Vehicles will be equipment with sensors to automatically detect people who cross the road or other cars that approach this car. The AI technologies will be used in developing automatic vehicle control systems to guide the automobiles based on the information collected from the sensors. The AI-controlled automobiles will be capable to go through the crowded intersections or roadways and avoid accidents, quicksand pits, and other possible cases where precaution must be taken.

Finally, AI will also have great potentials in public transportation management. AI techniques can be applied in real-time bus routing and scheduling and real-time passenger information system. The AI-based management system will provide the operators with more powerful and flexible decision-making tools based on the real-time information, so that improve the quality and security of the public transportation service.

KRISTEN L. SANFORD-BERNHARDT
Lafayette College

Many transportation researchers have eagerly embraced AI and advanced computing methods because of the potential of these methods to deal with seemingly intractable problems in the field. Transportation systems are complex systems, involving many interacting components and actors. Transportation problems not only involve substantial amounts of uncertainty in measured data; those that go beyond physical phenomena involve people. This is one of the things that make transportation problems so interesting, of course, but it also makes them difficult to model. As computational techniques and computing power have evolved, transportation engineers and planners have been able to improve their methods. To me, this increasing ability to model complex systems is one of the greatest promises of AI and advanced computing.

NICOLAS SAUNIER
École Polytechnique de Montréal

This document is a testament to the contributions of AI and advanced computing techniques to transportation and there is every reason to think that this will continue. In particular, systems relying on these techniques should become more and more autonomous. As authors have noted repeatedly in this circular, a major disadvantage of current techniques is the selection of parameters and the lack of guidance, which leads to painstaking and time-consuming manual trial and error and sub-optimal operation. Truly autonomous systems will minimize user input and supervision by being able to adapt automatically to various environments and conditions.

A perhaps overlooked and significant force in this progress is the Internet and its enabling power; it enables, in particular, at a minimal cost the sharing of knowledge and collaboration in spatially distributed network of individuals. The use of AI and advanced computing techniques has become more widespread because many tools are freely available, many under open source licenses, because they can be re-used and modified which allows a better and faster bootstrapping of technology. Demonstrating the principles of swarm intelligence, re-use, sharing, and collaboration allow human society to become more intelligent and better equipped to adapt

to a changing environment and to cope with the complexity and challenges of ensuring safe and fast transportation for a growing population.

GARY S. SPRING
Merrimack College

When considering advancements in any field, it is instructive to look back as we move forward. AI applications in the transportation field began, about 25 years ago, with a main focus on expert systems software. The field has fairly exploded since then. As new AI technologies have emerged over these two decades, many follow the path of the expert systems and become part of our mainstream tool box. Some areas that I believe hold the most promise in the coming decades are:

- Learning systems that allow for continuous improvement in performance.
- Agent-based systems along with other advanced simulation approaches that allow for short- and long-term prediction of traffic and traffic behavior in real time.
- Communications and information processing technologies that will someday allow for collection of traffic data in real time.
- Heuristic systems that allow for decisions to be made that are based upon imprecise information. These include fuzzy systems that are able to accommodate the uncertainties of real operational systems make possible fully automatic controls
- Better and smarter analysis tools may augment or even replace traditional tools for assessing and predicting safety and efficiency; for example, reinforcement learning as a replacement for Bayesian statistics; and cellular automata as a replacement for simulations based upon underlying probability distributions.

In these past couple of decades we have seen some changes in transportation but given the imperatives of providing sustainable systems, an improved infrastructure, lower congestion on our highways, and a safer driving experience, it is certain that the future will necessarily bring significant changes in transportation. AI and advanced computing will fuel these changes.

RAMKUMAR VENKATANARAYANA
Southeast Vipassana Center

The transportation field, across all modes and at every stage of the life cycle of every process, is being inundated with more and more data collection. This has been the natural evolution, given the increasing calls for transparency and accountability from the public, in addition to the long-existing demands for better efficiency, safety, and environmental consciousness. However the uncertainties with the data as well as the (planning or prediction) models remain. I think that extracting clear, timely, and useful knowledge from much larger datasets than in the past will be the mandate of the future. There is also an increasing trend for the public to demand real-time and accurate predictive information. To cater to all these demands, nonlinear approaches to data classification, pattern or knowledge discovery, data compression, data storage, analysis and prediction, and visualization will be essential. As the prospects loom for us to witness and enjoy

the benefits of fully automated vehicles, I believe the need for both research and application of AI methods and algorithms will see a rise in the future. A relevant example is the new Jeopardy champion, Watson, a super computer designed and developed by IBM. Watson is very knowledgeable and adaptive to learn and answer a myriad and very large number of questions. A much more knowledgeable, flexible, agile, confident, as well as empathetic, computer is necessary to drive a vehicle down the road. Several experiments are currently being carried out across the world, for realizing this vision. I predict that we will definitely need out-of-the-box thinking, and AI algorithms, to get there.

BILLY M. WILLIAMS
North Carolina State University

The many AI methods presented in this circular hold much promise as additions (and possible additions) to the toolbox for transportation system planning, design, and operation. Along with the promise, there is also risk. Many of the techniques described are powerful and robust at giving good answers (or at least what appear to be good answers) easily and for a broad array of applications. Some of the techniques in essence are, and many have the potential to be, black boxes in the fullest sense. In order for the transportation research and practice communities to realize the promise while avoiding the pitfalls, we must be diligent to make sure that we deepen our knowledge at least as fast as we develop new tools. Careful readers of this circular will have noticed multiple time where the authors have extolled various AI techniques for being simple to implement. Having powerful and simple techniques brings the temptation to uncritical, indiscriminant use. However, for the dedicated transportation professional the availability of such tools brings the responsibility to understand the techniques well and apply them wisely. Richard Hamming famously said that "the purpose of computing is insight, not numbers." Hamming was a mathematician. So his attention was more finely focused on knowledge acquisition, i.e., insight, than is ours as transportation problem solvers. However, our purpose is not merely to find a way to solve each particular problem. We, too, must be unrelenting in our efforts to expand our insight so that we can work together to provide a transportation system that is increasingly safe and efficient. The ultimate value of AI techniques applied to transportation will be directly related to the real intelligence of those who develop and implement them.

YUNLONG ZHANG
Texas A&M University

Transportation systems in general are complex. The analysis and modeling of a transportation system often involves a multitude of variables and complex relationships among them. In many cases, these relationships are difficult to identify and cannot be accurately molded by traditional approaches. The analyses of many areas and elements in transportation often have to deal with the issues of missing, inaccurate, or erroneous data, which often make traditional methods inapplicable. Conventional methods also lack the powers to incorporate uncertainty and fuzziness, which have become more and more important from the perspectives of both theoretical and practical considerations. Many decision-making processes in transportation require human reasoning, learning from past experience, and adapt to new and future scenarios.

Because of the ability of humanlike thinking, AI approaches can be developed to aid decision making processes involving large number of variables, complicated relationships, and incomplete data that are common in the field of transportation. For this reason, AI has its unique advantages over some traditional analytical procedures in transportation modeling, analysis, control, and management. AI methods can be used to obtain solutions where traditional methods would fail or cannot be applied. AI can also provide alternative solutions for some modeling problem.

Other than use of AI to model and analyze transportation data, predict system performance, and aid in decision making for management and control, another promising field of application of AI in transportation is the applications of robotics. Robots can potentially be developed and deployed to conduct duties as heavy construction and in areas such as hazardous sites and locations.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

TRANSPORTATION RESEARCH BOARD
500 Fifth Street, NW
Washington, DC 20001