

Guest Editorial: Big Data for Driver, Vehicle, and System Control in ITS

DATA collection has long been considered to be part of the daily routine for traffic engineers and transportation planners to support traffic control, transportation planning, traveler information systems, and transport system performance evaluation. Conventionally, traffic data are primarily obtained from fixed-point traffic surveillance systems, such as loop detectors (e.g., single or double loops), roadside sensors (e.g., microwave vehicle detection system), or overhead detectors [e.g., laser-based detection systems (LBDS)]. Studies have been conducted to use buses, taxis, or commercial vehicles as probes to collect vehicle positioning and speed data for travel time or traffic state estimation. In recent years, with advancements in new technologies, especially the Internet, huge volumes of traffic or traffic-related data can now be made available to us from a variety of sources on a real-time basis. Big data in transportation, in the form of the quickly moving data stream, are typically full of noise. Moreover, the information embedded in those data can quickly become obsolete. The ability to capture, process, and analyze these data and turn them into valuable information in a timely manner for intelligent transportation systems (ITS) has posed major challenges that have to be addressed by both researchers and practitioners. This special issue covers some of the recent research on big data for drivers, vehicles, and system control in ITS. A total of 11 papers are included in this special issue.

The collection of papers included can be categorized in several ways. In terms of applications, the special issue covers transit network design, taxi and bus operation, traffic flow and travel time prediction, dilemma zone management, and geometric design of roads. In terms of data sources, it covers both conventional sources of data such as bus, taxi, and individual vehicle trajectories, and also novel sources of data such as those based on the connected vehicle technology, roadside communication equipment, or onboard devices. In terms of techniques or methodologies developed for processing big data, this collection includes papers covering the data imputation technique based on the k -nearest neighbor method, data management infrastructure, the machine learning approach, and the graph-theory-based method.

The first paper, “TORD Problem and Its Solution Based on Big Trajectories Data,” by Li *et al.*, utilizes GPS-equipped mobile devices to generate vehicle trajectory data for identifying preferable routes by defining the problem as a temporal-sensitive optimal route discovery problem. A two-layer preference network is constructed for mining the associations of historical trajectories. In addition, three approximation

algorithms are developed to solve the problem. Experimental results demonstrate the efficiency and effectiveness of the proposed algorithms.

The second paper, “Managing Spatial Graph Dependencies in Large Volumes of Traffic Data for Travel-Time Prediction,” by Salamanis *et al.*, presents a graph-theory-based technique for managing the spatial correlation of traffic between roads of the same urban traffic network. A STARIMA-based traffic-prediction model is proposed in their work, which takes into account the identified dependencies. The proposed technique is benchmarked using traffic data from Berlin (Germany) and Thessaloniki (Greece). Benchmark results indicate not only significant improvement on the computational time required for calculating traffic correlations, but also reveal that a different variant works better in different network topologies, after comparison to third party approaches.

The third paper, “Finding the Shortest Path in Stochastic Vehicle Routing: A Cardinality Minimization Approach,” by Cao *et al.*, focuses on finding an optimal path that maximizes the probability of on-time arrival at a destination. A data-driven approach, which directly explores the traffic-related big data, is proposed in the paper. In their approach, Cao *et al.* first reformulate the original problem as a cardinality minimization problem directly based on the travel time sampled from each road link. They then apply an l_1 -norm minimization technique to solve the reformulated problem. This problem is subsequently transformed into a mixed integer linear programming problem, which can be solved using standard solvers.

This special issue also includes a paper on taxi mode that, together with other emerging competing forms of taxi-like modes (e.g., Uber), plays an increasingly important role in improving mobility of people. The fourth paper is entitled “An Unlicensed Taxi Identification Model Based on Big Data Analysis,” by Yuan *et al.* In this paper, a smart model is proposed to identify unlicensed taxis. The model consists of two sub-model components, namely a candidate-selection model and candidate-refined model. The former is used to screen out a coarse-grained suspected unlicensed taxi candidate list. The list is taken as an input for the candidate-refined model, which is based on machine learning to get a fine-grained list of suspected unlicensed taxis. The proposed model is evaluated using the real-life data.

The fifth paper, “An Integrated Dilemma Zone Protection System Using Connected Vehicle Technology,” by Zha *et al.*, proposes an evaluation framework for dilemma zone protection via vehicle-to-infrastructure. The proposed framework utilizes data obtained from roadside equipment for actuated signal operations and dilemma zone protection. They performed a simulation study. The simulation results indicated

that, with a 40% or higher penetration rate, the proposed framework provides the best result for dilemma zone protection that would also result in a reduction in delay from conflicting movements.

The sixth paper, “Data-Driven Transit Network Design From Mobile Phone Trajectories,” by Pinelli *et al.*, presents a data-driven method for transit network design that relies on a large sample of user location data that is available from mobile phone telecommunication networks. The method works by deriving frequent patterns of movements from anonymized mobile phone location data and merging them to generate candidate route designs. Additional routines for optimal route selection and service frequency setting are then employed to select a network configuration made up of routes that maximize system-wide traveler utility.

The seventh paper, “On Geocasting over Urban Bus-Based Networks by Mining Trajectories,” by Zhang *et al.*, employs bus trajectory data to build a probabilistic spatial-temporal graph model that provides the routing paths with the best possible quality of service levels for data delivery requests. Taking advantage of the distinctive features of city bus networks such as area-wide coverage and fixed bus routes, they demonstrate the potential of forming the communication backbone in vehicular ad hoc networks (VANETs). In their work, they primarily focus on the geocast in the bus-based VANETs and present a geocast routing mechanism called Vela, which analyzes and mines historical bus trajectories and characterizes the spatial-temporal traffic patterns. The results of experiments on the real and synthetic trajectories show that Vela performs much better in terms of delivery ratio and delay.

The eighth paper, “A Flexible System Architecture for Acquisition and Storage of Naturalistic Driving Data,” by Bender *et al.*, describes the design principles behind data management infrastructure, with an emphasis on flexibility and maintainability. This is achieved by breaking up codes into a modular design that can be run on many independent processors. Messages passing over a publish-subscribe network enable interprocess communication and promote data-driven execution. By following these principles, rapid prototyping and experimentation with new sensing modalities and algorithms are possible. Furthermore, they compare the communication library underpinning the proposed architecture against several popular communication libraries. Features designed into the system make it decentralized, robust to failure, and amenable to scaling across multiple machines with minimal configuration. Code written using the proposed architecture is compact, transparent, and easy to maintain. Experimentation shows that the proposed architecture is high performing when compared against alternative communication libraries.

The ninth paper, “Data-Driven Imputation Method for Traffic Data in Sectional Units of Road Links,” by Tak *et al.*, proposes a data-driven imputation method based on the concept of the k -nearest neighbors method. The proposed imputation algorithm is different from the conventional algorithms in three

aspects: 1) it can be implemented with incomplete historical data; 2) it applies a weighting matrix for calculating the similarity of k -nearest neighbors to reflect the spatial and temporal correlation of the transportation data; and 3) it can be processed under the distributed computing environment by segmenting the highway network into several sections. The authors compare the proposed imputation algorithm to the nearest historical data and bootstrap-based expectation maximization, by varying the missing data type, missing ratio, traffic state, and the day type. The results show that the proposed algorithm achieves better performance in almost all of the scenarios.

The tenth paper, “Using Analytics in the Implementation of Vertical and Horizontal Curvature in Route Calculation,” by Svenson *et al.*, deals with computing the vertical and horizontal curvature in a route, since the best route for logging trucks is difficult to determine as many road features need to be considered, including the vertical and horizontal curvature of a route. They propose a methodology for computing the perception of these features. The proposed methodology also includes the processes to clean and complement inaccurate coordinates in the data. The result of their work is implemented in the intelligent transport planning system, called calibrated route finder, which is used by the Swedish forestry agency.

The final paper, “Repeatability and Similarity of Freeway Traffic Flow and Long-Term Prediction Under Big Data,” by Hou and Li, proposes a long-term traffic flow forecasting model and hybrid forecasting algorithms for short-/long-term traffic flow prediction. In their work, they defined the similarity and repeatability measures for traffic flow patterns and examine the properties associated with these measures. Field data of 85 days are used to assess the performance of these measures.

WEI-HUA LIN

Department of Systems and Industrial Engineering
University of Arizona
Tucson, AZ 85721 USA
E-mail: whlin@email.arizona.edu

HONGCHAO LIU

Department of Civil Engineering
Texas Tech University
Lubbock, TX 79409 USA
E-mail: hongchao.liu@ttu.edu

HONG K. LO

Department of Civil Engineering and
Environmental Engineering
Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong, China
E-mail: cehklo@ust.hk



Wei-Hua Lin received the Ph.D. degree in civil engineering from University of California at Berkeley, Berkeley, CA, USA.

He is currently an Associate Professor with the Department of Systems and Industrial Engineering, University of Arizona, Tucson, AZ, USA. He was a Postdoctoral Researcher in the PATH program, University of California at Berkeley. His research interests include optimization in intelligent transportation systems, logistics, and transportation network analysis.

Dr. Lin is a past member of the Intelligent Transportation Systems Committee of the Transportation Research Board, National Research Council of the United States.



Hongchao Liu (M'08) received the Ph.D. degree in civil engineering from University of Tokyo, Tokyo, Japan.

He is currently a Professor with the Department of Civil, Environmental and Construction Engineering, Texas Tech University (TTU), Lubbock, TX, USA. Prior to joining the faculty of TTU, he was a Postdoctoral Researcher in the PATH program, University of California at Berkeley, Berkeley, CA, USA. His research interests include traffic operation and control, intelligent transportation systems, data analytics, and transportation system modeling.

Dr. Liu is a member of the Artificial Intelligence and Advanced Computing Applications Committee, Transportation Research Board, National Research Council of the United States.



Hong K. Lo received the B.Sc. degree in civil engineering from University of Hong Kong, Pokfulam, Hong Kong, and the Ph.D. degree in civil engineering from Ohio State University, Columbus, OH, USA.

He is currently the Head and a Chair Professor of civil and environmental engineering with The Hong Kong University of Science and Technology, Kowloon, Hong Kong, China. Prior to this, he was with University of California at Berkeley, Berkeley, CA, USA. He specializes in transportation system modeling, dynamic traffic assignment, traffic control, public transportation, and reliability analysis.