Introduction to the Special Issue on Online Learning for Big-Data Driven Transportation and Mobility

IN THE last years, the arrival and progressively gained maturity of technological paradigms such as Connected Vehicles, the Internet of Things, Sensor Networks, Urban Computing, Smart Cities, Cloud Computing, Edge Computing, Big Data and others alike have ignited the role historically played by data-based learning techniques to levels never seen before. This sharp increase has been particularly noticed in the design and management of intelligent systems for transportation and mobility, as processes, services and applications deployed in these systems are fed with data substrates captured at unprecedented rates and scales. Legacy sensing equipment installed on the roads' infrastructure (e.g. induction loops and cameras) are nowadays complemented by alternative means to sense the transportation and mobility context of interest in real time and ubiquitously, as can be exemplified by data collected in a crowd-sourced way by using ad-hoc smart applications, as well as floating car data and/or social media.

Interestingly, the diversity, speed and/or volume of data streams generated in transportation and mobility environments pose true Big Data challenges in terms of i) diversity/variety of such data flows, ii) the aggregate volume of the collected information, and iii) the high speed and non-stationarity of data streams. If proves worthy to develop strategies that learn from this complex data substrate efficiently and adaptively. It makes sense to develop agile methods to build always up-to-date models in real time and CPU/Memory contention. Such research trend has not only capitalized on the design of incremental ways to train data mining models so as mine data at the high speed at which they are generated (online learning), but also has placed special emphasis on how models should be adapted to changes in the distribution modeling the data.

This special issue has gravitated on this research avenue by gathering contributions dealing with efficient solutions to process and learn from ITS and mobilty data that feature the aforementioned characteristics. The ultimate aim of the special issue is to stress on the increasing relevance of considering, when designing data-driven processing pipelines, not only its performance to accomplish the task at hand, but also contextual factors such as their complexity, computational resource constraints or scalability. To this end, the special issue has congregated a number of high-level contributions showcasing this advocated call for looking beyond performance in data-based modeling towards ensuring their *accionability* within the ITS and Mobility realms.

In response to our open call for contributions, several submissions were received at the editorial office of IEEE Transactions on Intelligent Transportation Systems. Such submissions were processed and evaluated in accordance to the reviewing standards and rigor of the journal, considering their alignment with the scope of the special issue, originality, technical soundness and presentation quality. As a result, 9 submissions were finally accepted for publication. We heretofore provide a summary of such accepted papers, for the reader to better appraise the coverage of topics of the special issue.

To begin with, mining of spatial trajectory data generated by vehicles and pedestrians is arguably one of the ITS research areas where the above Big-Data requirements occur to prevail most, specially in terms of speed and scales. This is the focus of several contributions published in the special issue, starting with the paper entitled "A Hybrid Solution for Real-Time Travel Mode Detection and Trip Purpose Prediction", by Elton F. de S.Soares et al. In this work, a solution for real-time travel mode detection and trip purpose prediction is proposed. The proposed approach relies on a feature extraction algorithm that discriminates the best predictors for subsequently applying supervised machine learning algorithms for classification. Resampling techniques and automatic hyper-parameter tuning methods were also utilized for selecting the best model choice. Experiments revealed that the best model was found to be Random Forest with upsampling, scoring accuracies of 88% for travel mode detection and 81% for trip purpose prediction. Traffic data was also central for the study reported in "TraLFM: Latent Factor Modeling of Traffic Trajectory Data", by Meng chen et al. In this work a model coined as TraLFM was proposed to allow for a better understanding of human mobility patterns. To this end, latent factor modeling was applied over traffic trajectory data to infer mobility patterns, predict next locations. Furthermore, personal and temporal features were also fed to the generative process of trajectories, which permits to determine time periods with similar mobility patterns. Extensive experiments were analyzed, concluding that the proposed TraLFM approach outperforms other state-of-the-art methods used for next location prediction.

Another data-based task that also emerges concurrently in the ITS literature is the unsupervised detection of anomalies in data, which may also fall below Big-Data computational constraints. This is the case targeted by "Anomaly Detection

Digital Object Identifier 10.1109/TITS.2019.2955548

in Road Networks Using Sliding-Window Tensor Factorization", by Ming Xu et al. This work elaborates on a taxonomy of possible anomalies occurring in traffic data based on deviations from different spatial or temporal patterns. To efficiently detect such multiplicity of anomalies, a sliding window tensor-based framework is proposed and validated over a massive real-time vehicle trajectory dataset. Evidences are given on the efficiency of the approach for online data updating.

When it comes to vehicular perception, efficient algorithmic means are also needed to cope with the enormous scales at which information from the vehicular surroundings is captured and processed. The study in "Depth Embedded Recurrent Predictive Parsing Network for Video Scenes", by Lingli Zhou et al, aimed precisely at this direction. In particular, the work proposed a novel depth embedded recurrent predictive parsing network (RPPNet) for video scene parsing (semantic image segmentation), with applications to autonomous driving and navigation. The proposed model uses binocular images as its input so as to capture dynamic information between video frames and thereby, ensure a coherent flow of temporal and spatial features for the learning task. Such improved features are fused together with those supplied by conventional image parsing models. Experiments discussed in this work shed light on the performance gains attained by the proposed RPPNet approach over other methods, especially for the detection and segmentation of classes with small areas in the scene under analysis (e.g. pole, pedestrians).

Public transport has revealed itself to spawn enormous opportunities for Big-Data modeling, yet not exempt from computational challenges as the ones targeted in this special issue. Two contributions to the special issue support this claim. The paper entitled "Estimation of Influence Distance of Bus Stops using Bus GPS data and Bus Stop Properties", by Ilgin Gokasar et al, analyzed the traffic behavior of buses in the surroundings of bus stops, taking also into account contextual properties of the location of the bus stop under study. The goal was to determine whether different traffic behavioral dynamics (reflected in the so-called influence distance) are induced by such contextual properties (e.g. number of lanes, location of the traffic lights). This hypothesis was validated over 12 distinct bus routes from the city of Istanbul (Turkey), comprising daily GPS data of more than 5000 bus vehicles and 450 bus stops. Public bus service was also spotted in "Standing Passenger Comfort: A New Scale for Evaluating the Real-time Driving Style of Bus Transit Services", by B. Barabino et al. This work advanced over previous objective and subjective driving style characterization studies performed in isolation with respecto to each other in the past. Subjective feedback from the standing passengers was integrated with objective measurements of longitudinal and transversal accelerations of the bus. The framework proposed in this study was put to practice on a real case study with data provided by an Italian bus operator. Results buttressed the capability of the framework to assess the On-Board Comfort Level (OBCL) in urban bus transit services in real-time.

Architectural contributions for the deployment of online learning techniques and concept drift adaptation methods also found their place in the special issue. The platform presented in "Online Incremental Machine Learning Platform for Big Data-Driven Smart Traffic Management", by Dinithi Nallaperuma et al, was designed to facilitate the use of unsupervised online incremental machine learning, deep learning, and deep reinforcement learning methods over dynamic, volatile Big-Data streams generated by traffic sensors. The platform was validated by performing descriptive (emotion analysis), predictive (traffic forecasting) and prescriptive (traffic control) tasks based on real-time Bluetooth sensor network data and social media data from an arterial road network in Victoria, Australia.

Finally, we pause at the two last works rounding up the special issue, both elaborating on the vast opportunities paved by Big-Data for car-sharing and electro-mobility. The first of these contributions, entitled "DeepPool: Distributed Model-free Algorithm for Ride-sharing using Deep Reinforcement Learning" by Abubakr Alabbasi et al, dealt with the cost-efficient allocation of vehicle capacities in ride-sharing platforms. Specifically, a distributed model-free Deep Reinforcement Learning approach (DeepPool) was proposed to learn optimal dispatch policies in dynamic ride-sharing environments, wherein the trip time and the actual routes vehicles should take were updated continuously and reflected on the learned policies. The proposed DeepPool model is designed adapt to changing distributions of customer arrivals, active cars, vehicles' locations, and users' valuations for the total travel time. Real taxi trip records in New York City (USA) were used to build a large-scale realistic simulator, on which to test the performance of DeepPool and assess its capability to reduce the number of operating vehicles and consequently, reduce the traffic congestion. The second work along this line is "Free Floating Electric Car Sharing: A Data Driven Approach for System Design", by Michele Cocca et al. In this study a data-driven approach is followed to build a realistic simulator for electric car sharing scenarios. To this end, actual rental traces are exploited to analyze, via simulations, the impact of the location of charging stations, and car return policies. This study serves as an illustrative example of the possibilities of Big-Data beyond modeling, as they can be also used to construct simulators featuring unprecedented levels of realism.

We sincerely hope that the reader enjoys this special issue just as much as we did. However, there still remain much to be done around Big-Data and online learning for ITS and Mobility. We are immersed in an exciting age where data-based modeling pops almost from every single problem or research line under study. It is our responsibility to pursue all research directions rooted on this paradigm, not only in terms of new models and learning methods, but also in all practical matters aimed at ensuring that data-based models lead to useful decisions. In short: actionability.

JAVIER DEL SER, *Guest Editor*
TECNALIA
48160 Derio, Spain
e-mail: javier.delser@tecnalia.com

JAVIER J. SANCHEZ-MEDINA, *Guest Editor*
Innovation Center for the Information Society
University of Las Palmas de Gran Canaria
35001 Las Palmas, Spain
e-mail: javier.sanchez.medina@ieee.org

ELENI I. VLAHOGIANNI, *Guest Editor*
Department of Transportation Planning and Engineering School of Civil Engineering
National Technical University of Athens
Zografou Campus
15773 Athens, Greece
e-mail: elenivl@central.ntua.gr

Javier Del Ser (M'07–SM'12) received the first Ph.D. degree *(cum laude)* in electrical engineering from the University of Navarra, Spain, in 2006, and the second Ph.D. degree *(cum laude)*, extraordinary PhD prize, in computational intelligence from the University of Alcalá, Spain, in 2013. He is currently a Research Professor in artificial intelligence and a Leading Scientist of the OPTIMA (optimization, modeling and analytics) research area with TECNALIA, Spain. He is also an Adjunct Professor with the University of the Basque Country (UPV/EHU), an invited Research Fellow with the Basque Center for Applied Mathematics (BCAM), and a Senior AI Advisor with the technological startup SHERPA.AI. He is also the Coordinator of the Joint Research Lab between TECNALIA, UPV/EHU, and BCAM, and the Director of the TECNALIA Chair in artificial intelligence implemented at the University of Granada, Spain. He has published more than 280 scientific articles, co-supervised ten Ph.D. theses, edited seven books, co-authored nine patents, and participated/led more than 40 research projects. His research interests are in the design and application of artificial intelligence methods for data mining and optimization for problems emerging in intelligent transportation systems, smart mobility and logistics, and among others. He was a recipient of the Bizkaia Talent prize for his research career. He has served as a Guest Editor in journals from the ITS field, such as the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS and the IEEE ITS Magazine. He is an Associate Editor of tier-one journals from areas related to data science artificial intelligence, such as *Information Fusion*, *Swarm and Evolutionary Computation*, and *Cognitive Computation*.

Javier J. Sanchez-Medina (M'11–SM'17) is currently an Associate Professor with ULPGC, Spain. He has extensive experience in the development of traffic models and simulation platforms. His research interests include mainly the application of data mining, machine learning, and parallel computing to intelligent transportation systems. He has published more than 60 publications in these areas. He is also a very active volunteer in the ITS community, serving regularly at IEEE conferences in their organizing committee, and as an IEEE ITS Society's Officer and a BoG member. He has also been a Keynote Speaker at ANT2017 and SOLI2017 and a Distinguished Lecturer at the IEEE ITSS Tunisian Chapter in 2016.

Eleni I. Vlahogianni (M'07) received the Diploma degree in civil engineering and the Ph.D. degree in traffic operations from the National Technical University of Athens (NTUA), Greece. She was a Visiting Scholar with the Institute of Transportation Studies, University of California, Berkeley, USA, working on the development of an Arterial Performance Measure System (APeMS). She is currently an Associate Professor with the Department of Transportation Planning and Engineering, National Technical University of Athens. She has a strong algorithmic background, which includes nonlinear dynamics, applied statistical modeling, data mining and machine learning techniques, and computational intelligence. Her professional and research experience includes projects and consultancies, in a national and European level focusing on urban mobility and traffic flow management, public transport, and traffic safety. Moreover, she is an author of more than 50 articles in international peer-reviewed journals and more than 60 publications in international conference proceedings with over 3000 citations. Her primary research field is traffic flow analysis and forecasting. Other research fields include mobility modeling, driving analytics, ICT applications to transportation, intelligent transportation systems, traffic management, and advanced technologies for monitoring transportation infrastructures. She is the Chair of the IEEE ITSS Technical Committee on Smart Cities and Smart Mobility. She is an Associate Editor of *Transportation Research Part C: Emerging Technologies*, the *IEEE ITS Magazine*, and *International Journal of Transportation Science and Technology*.