Computational Intelligence and Optimization for Transportation Big Data: Challenges and Opportunities

Eleni I. Vlahogianni

Abstract With the overwhelming amount of transportation data being gathered worldwide, Intelligent Transportation Systems (ITS) are faced with several modeling challenges. New modeling paradigms based on Computational Intelligence (CI) that take advantage of the advent of big datasets have been systematically proposed in literature. Transportation optimization problems form a research field that has systematically benefited from CI. Nevertheless, when it comes to big data applications, research is still at an early stage. This work attempts to review the unique opportunities provided by ITS and big data and discuss the emerging approaches for transportation modeling. The literature dedicated to big data transportation applications related to CI and optimization is reviewed. Finally, the challenges and emerging opportunities for researchers working with such approaches are also acknowledged and discussed.

1 Introduction

With a vast number of diverse Intelligent Transportation Systems (ITS) operating Worldwide, web-based, mobile, and sensor generated data arrive at and overwhelming scale. This availability allows for new science paradigms to be introduced and novel insights to be gained. Traditionally, turning data into knowledge relies on classical statistical analysis and interpretation; this fundamentally requires analysts to become intimately familiar with the data and serve as an interface between the data and the users. With the recent availability of very large data sets (big data), this form of manual probing becomes slow, expensive, and frequently unfeasible. Methodologically, new approaches are needed to efficiently deal with some of the challenging issues related to big data; some of them are data size, high dimensionality, overfitting,

E.I. Vlahogianni (✉)
Department of Transportation Planning and Engineering, School of Civil Engineering,
National Technical University of Athens, Athens, Greece
e-mail: elenivl@central.ntua.gr

N.D. Lagaros and M. Papadrakakis (eds.), *Engineering and Applied Sciences Optimization*, Computational Methods in Applied Sciences 38,
DOI 10.1007/978-3-319-18320-6_7

assessing statistical significance, rapidly changing, missing and noisy data, complex relationships between fields, user interaction and prior knowledge, and system integration.

Big Data is growing exponentially due to the growth of both existing and new data sources (e.g. geospatial, social media comments, mobile). To build a smarter planet, we need smarter computing—computing that is tuned to operate, managed through the cloud and, importantly, designed for big data. Novel modeling paradigms will have to: i. Capture and manage high volume multi-source data encompassing text, images, sounds, generated impulses etc. ii. Understand patterns unfolding in time across a complex transportation system (spatial unfolding) and produce critical information and alerts.

In this context, Computational Intelligence (CI) offers an excellent alternative to traditional hypothesis-driven (i.e. deductive) statistical data analyses and attempts to extract meaningful patterns in big data. In Transportation, there has been increased interest among both researchers and practitioners in exploring the feasibility of CI algorithms in transportation problems, especially related to optimization. The advantage of CI data analysis applications over other alternatives lies in their flexibility, their ability to discover unknown mechanisms and covariations elusive to statistical approaches, their accuracy, and their ability to handle dynamically changing big data. Still, the development of efficient CI applications in Transportation is complex, rarely taught in transportation programs in Academia, while model development and validation are frequently done ad hoc and do not follow universally accepted procedures.

In this paper, the unique opportunities created by the data obtained from modern ITS are discussed and some of the emerging approaches for handling big data are reviewed. The literature dedicated to big data transportation applications related to CI and optimization is reviewed. Finally, the challenges and emerging opportunities for researchers working with such approaches are also acknowledged and discussed.

2 The "New" Transportation Landscape

Urbanization, smart cities and disruptive technologies may be considered as the three pillars transforming the transportation arena. Urban areas are, nowadays, considered as the dominant type of settlement for humanity. In this context, optimizing transportation and mobility play an imperative role in sustainable urban development. Second, cities are becoming smarter, in terms of their infrastructure, with the aim to maximize resources and actively support sustainable growth and high quality of life, through participatory action and engagement, while preserving natural resources [18].

To be able to fully benefit of the above, a transportation system should be instrumented, interconnected and intelligent. In this context, there is an increasing interest in finding novel technologies to support the transportation arena. Some of the most prominent are mobile communications, cloud technologies, energy storage, autonomous vehicles and the Internet of Things (IoT). The latter is a novel concept straightforwardly applicable to transportation applications; IoT consists of a variety of devices or objects—such as Radio-Frequency IDentification (RFID) tags, sensors, actuators, mobile phones, and so on—which, through unique addressing schemes, are able to interact with each other and cooperate with their neighbors to reach common goals [4, 106]. By continuously collecting, analyzing and redistributing transportation information, IoT networks can offer valuable, real time information to both travelers and operators, and, thus, support and improve the operations of ITS, traffic and public transportation systems.

3 Big Data and Transportation

3.1 A Definition

Most widely available definitions of "big data" converge to the following: any collection of data is big or may become big, when it becomes difficult or impossible to model its complexity using traditional data processing tools. This definition leave much room for arguments and misconceptions about what data can be considered as big and how big are the available data.

A more scrutinized look at big data introduces the concept of three V's: big data are quantities amounts (Volume), of any type (Variety), that are collected at unprecedented speed and must be dealt with in a timely manner (Velocity) [71]. The V's can be extended to include acyclic or irregular temporal data (Variability), the uncertainty stemming from the difficulty in controlling the quality and accuracy of the data (Veracity).

3.2 Sources and Applications of Big Data in Transportation

The big data phenomenon is not new in Transportation and Traffic Engineering. The leading edge of transportation data has for long been streaming data coming for a variety of sensors (loop detectors, video cameras, weather stations etc.). What has changed over the years is the cost of new monitoring systems (more economic ways of producing streaming data, such as the passive data produced by personal GPS), the data granularity (very detailed information collected in real time) and the availability of new sources of unstructured or semi-structured data, such as logs, clickstreams, and social media data (tweets, Facebook posts etc.). A detailed classification of Big Data sources may be found in Hashem et al. [50].

The intrusion of big data and analytics to the transportation research and industry is significant. Large companies including Google, IBM, SAS, INRIX etc. systematically fund research and applications on how to leverage big data of all forms (structured and unstructured) to improve transportation services and customer satisfaction, manage transportation infrastructure, as well as predict or estimate traffic conditions. The gains from using big data in transportation are numerous for road users, authorities and private sector. Road users can make informed decisions to save time and reduce their personal trip cost based on continuously available traffic information from various sources of the road network with extended spatio-temporal coverage. Road authorities may take advantage of big data to understand travel patterns to identify policy interventions, control traffic and manage demand and congestion, or even change the users' behavior. Finally, private sector may gain significant competitive advantage by identifying prevailing trends or increase productivity by improving their route planning and logistics.

A field that has profited the most from the advent of big data is travel demand estimation; various approaches to derive OD matrix and mobility patterns have been based on mobile phone and personal GPS data [16, 42, 61, 74, 76, 88]. Papinski et al. [89] and Bierlaire et al. [13] developed a route choice behavior based on personal GPS traces, whereas Hood et al. [53] used GPS traces to develop a bicycle route choice model. Liu et al. [76, 79] studied land uses based by analyzing GPS-enabled taxi data in Shanghai. Cai et al. [17] analyzed the manner travel patterns may influence the electric vehicle charging infrastructure development using trajectory data from taxis in Beijing. Chen and Chen [24] utilized taxi GPS traces for nigh bus routes planning.

Regarding traffic, mobile phone counts have been systematically used for extracting traffic information in the form of volume, speed and density in both urban and suburban road networks [3, 9, 10, 51]. Castro et al. [19] used taxi GPS traces to estimate the traffic flow conditions in urban areas. Guido et al. [47] attempted to infer speeds using GPS smartphone traffic probes.

Location based services and social media are the new hype for collecting transportation related data. Cheng et al. [27] and Cheng et al. [28] addressed issues of urban mobility by analyzing twitter and social networking data. Collins et al. [33] proposed a sentiment analysis approach to measure transit rider satisfaction by quantifying twitter feeds. Hasan and Ukkusuri [49] demonstrated the use of a large-scale geo-location data set to analyze and understand individual activity patterns. Recently, Yang et al. [122] analyzed Foursquare data to derive OD information for non-commuting trips.

A new field of research that emerged from gathering individual data collection—either through smartphones or instrumented vehicles—is the extraction of driver's profiles during driving [5, 83, 84, 95, 100, 104, 110, 115, 119]. The scope of such profiles is to improve the efficiency during driving and mitigate risky behaviors that may lead to near misses or crashes. Driving big data has also been systematically used to develop advanced insurance systems based on the time and manner a user drives (pay as you drive, pay how you drive) [6, 85, 86].

4 Transportation Big Data Analytics

Analyses based on data, regardless of being big or not, have been recognized as a valuable tool for transportation operations. The stake when using big data is to be able to transform data into knowledge. Transforming data into knowledge involved a set of processes that are described in Fig. 1.

Each step towards the ultimate goal involves a set of tasks. For example data capturing and management involves indexing, searching, querying and visualization. The analysis stage may target to detect anomalies, reveal patterns and complex relationships. The prediction step entails complex and flexible data driven models that may consistently and accurately provide information on the future conditions, whereas mechanisms to create and disseminate information are the final step.

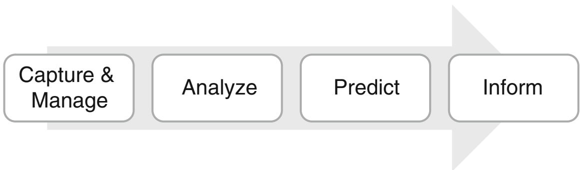
From a modeling standpoint, the problem faced with big data are numerous; first, these datasets are frequently of high dimensionality, meaning that they are difficult to visualize and understand. Moreover, having and extended dataset may not always mean having a representative dataset or a dataset with "perfect" information. The latter signifies that there is a need for a powerful preprocessing stage to assure that the models developed may be estimated and generalize real world conditions. Finally, assessing the statistical fit in big multi-dimensional datasets is not an easy task. Even when using data driven models, the surplus of data may lead to overfitting and models with reduced generalization power.

4.1 From Statistics to Computationally Intelligent Models

Usually, the statistical tools implemented entail several structural constraints and are unable to work on quirky and messy data with little or no structure. The lack of diversified statistical tools for big data analyses lead statisticians to see big data as a burdensome rather than a source of valuable information. A typical example is the time series of road traffic characteristics; typical autoregressive statistical models suppress or ignore nonlinearity and irregularities, whereas literature has systematically underlined the usefulness of these irregularities to understand the transitional nature of traffic flow [64, 107, 110, 111, 113, 114].

Evidently, with the advent of multi-source data collection systems, transportation datasets will not become perfect. Treating big data brings forward the focus on size,

Fig. 1 Processes in big data analytics



the ability to model messiness and multi-dimensionality in datasets, as well as the importance of correlations along with causation; we do not have to always understand the underlying mechanisms of the data to make them work to our benefit. To this end, new flexible and powerful modeling paradigms are imperative that are robust to imperfections and hypothesis free. The need to develop new analysis paradigms for the rapidly growing datasets has been underlined since the late 90s' [39]. This road contains either new forms of statistical thinking or data mining and computational intelligent models. Computational intelligence (CI) is the new hype in transportation modeling. CI includes neural networks, fuzzy logic, swarm intelligence, evolutionary algorithms, expert systems, agent based modeling etc. These models are applicable to many data mining problems, from warehousing to prediction and decision making, and may be proven more efficient due to their non-parametric hypothesis free nature.

Contrary to common thinking, some CI tools may bare significant similarities to classical statistical models, an issue frequently disregarded by connectionists that are more interested in producing accurate results rather than judging on the quality of their models and the properties of the error [15]. With the use of statistical inference, researchers may construct CI models equivalent to many popular statistical models [66]. For example, a single Perceptron is a linear regression model [93], while a Multilayer Perceptron with one hidden unit and a logistic function at the output layer is equivalent to a logit model [107].

The importance of CI to transportation is significant; CI may be used to develop scalable, manageable, adaptable and affordable transportation systems using common sense reasoning, perception and learning, as well as autonomy. One of the many advantages of CI, which is among the main differences with statistical thinking, is the ability of the latter to treat many "non-algorithmizable" problems (natural language processing, visual perception, character recognition etc.). Their ability to augment or replace human skills reflects to gains in computations, accelerates processing and increases productivity. These features may lead to providing results with improved accuracy and quality in a timely manner.

5 Computational Intelligent Optimization for Big Data Problems

In the entire process of mining knowledge from data, several modeling stages may be formulated as optimization problems. Optimization targets the "optimum" solution(s) for a given problem within allowable time. The issue is that each problem may have several local optimal solutions. The difficulty in converging relates to the problem's dimension and the number of objectives (large-scale multi-objective optimization). Evidently, large-scale optimization processes are affected by the curse of dimensionality in numerous ways [29]; the larger the dimensions of the phenomenon, the larger the solution space will be. The larger the dimension of a problem, the greater the risk of some problem characteristics to be altered with the scale.

Moreover, most traditional methods can only be applied to continuous and differentiable functions. Nevertheless, these conditions do not hold for most real world. The above complexities may be treated by problem decomposition strategies, surrogate-based fitness evaluations, data transformations etc. [58]. Another issue that may increase the complexity of the optimization problems is the spatio-temporal evolution of the datasets. In non-stationary environments and transportation problems (e.g. traffic flow) the dynamics may impose different optimal solutions in relation to time and space. This means that an optimization strategy should be able to treat dynamic problems and continuously converge to a solution.

CI approaches have both the structural flexibility and learning capability to deal with complex, time varying multi-objective problems [128]. CI applications to transportation include nature-inspired algorithms (evolutionary algorithms, particle swarm optimization etc.) and non-linear mapping and knowledge embedding approaches (neural networks, fuzzy algorithms etc.). CI have been found to perform well in non-stationary and highly nonlinear problems due to their robustness (impose little or no requirements on the objective function) and flexibility to handle highly non-linear mappings [54]. Moreover, self-adaptation and parallel operation are among the most important characteristics that enable CI to improve their performance and decompose complex tasks into simpler ones. Nevertheless, literature systematically underlines the need to cautiously apply CI to transportation problems as their proper development is frequently tedious and involves significant parametrization [66].

5.1 Computational Intelligent Optimization in Transportation Problems

Numerous efforts dedicated to CI optimization approaches to transportation applications can be traced in literature. Table 1 is a non-exhaustive list of the most recent research attempts related to CI and optimization. These applications are categorized by the transportation problem they aim to solve, the CI algorithms implemented, as well as the type of data used to evaluate the proposed approach. Special attention is given to whether the listed applications involve the full big data perspective (5 Vs).

Genetic algorithms may be considered the first and leading CI techniques in transportation optimization problems systematically applied to network design problems [67], vehicle routing and allocation problems [2, 44, 65, 78], signalization optimization [21, 22, 91, 99, 101] and highway alignment optimization [55, 63], pricing [68] and so on.

Significant interest from transportation modelers has been placed on Swarm Intelligence (SI). SI is an innovative branch of meta-heuristics derived from imitating the behavioral pattern of natural insects. Teodorović [102] reviews the literature on swarm intelligence and transportation and traffic engineering applications, whereas Zhang et al. [127] conduct a thorough review on the swarm intelligence applications to transportation logistics.

Table 1 Classification of literature on computational intelligent application to transportation optimization problems

Authors	Date	Problem	CI method	Data
Bai et al. [7]	2014	Transportation asset management	NSGA II	Numerical example
Chen et al. [25, 26]	2014	Trip planning	Heuristic Algorithm	Location-based social network, taxi GPS digital footprints[a]
Chira et al. [31]	2014	Vehicle routing	Evolutionary algorithms, ant colony	Real world case study
Danalet et al. [36]	2014	Pedestrian routing	Bayesian networks	Wi-fi data[a]
Doolan and Muntean [37]	2014	Vehicle routing	Ant-colony optimization	Simulation
Fagnant and Kockelman [38]	2014	Share autonomous vehicles	Agent-based model	Simulation
Forcael et al. [40]	2014	Tsunami evacuation routes	Ant colony	Real world case study
Galland et al. [43]	2014	Car pooling	Agent-based model	Simulation
Kallioras et al. [59]	2014	Emergency inspection scheduling	Harmony search	Real world case study
Kammoun et al. [60]	2014	Traffic routing	Ant-hierarchical fuzzy model	Simulation
Lin and Ku [75]	2014	Stopping patterns for passenger rail transportation	Genetic algorithm	Real world case study
Liu et al. [78]	2014	Emergency medical service allocation	Genetic algorithms	Real world case study
Pahlavani and Delavar [87]	2014	Route planning	Weed colonization	Simulation
Stolfi and Alba [98]	2014	Traffic routing	Evolutionary algorithm	Simulation
Terzi and Serin [103]	2014	Maintenance works on pavements	Ant colony	Numerical example
Yang et al. [122, 123]	2014	Highway alignment optimization	Genetic algorithm	Real world case study
Yin et al. [124]	2014	Hurricane evacuation	Agent-based model	Simulation
Zhang et al. [125]	2014	Transit network design	Agent-based model	Simulation
Zhou et al. [128]	2014	Mobile traffic sensor routing	Ant colony, PSO	Simulation
Arango et al. [2]	2013	Berth allocation	Genetic algorithms	Simulation

(continued)

Table 1 (continued)

Authors	Date	Problem	CI method	Data
Chevrier et al. [30]	2013	Railway scheduling	Evolutionary algorithm	Real world case study
Cong et al. [34]	2013	Traffic routing	Ant colony algorithm	Simulation
Goksal et al. [46]	2013	Vehicle routing	PSO algorithm	Numerical example
Jia et al. [56]	2013	Transportation-distribution planning	NSGA II algorithm	Numerical example
Kontou et al. [69]	2013	Transit depot allocation	Genetic algorithm	Real world case study
Lagaros et al. [70]	2013	Fund allocation	PSO algorithm	Real world case study
Levin and Kanza [73]	2013	Vehicle routing	Heuristic algorithm	Location-based network[a]
Liu et al. [77]	2013	Freeway corridor diversion control	Genetic algorithms	Real world case study
Shafahi and Bagherian [94]	2013	Highway alignment optimization	PSO algorithm	Numerical example
Ceylan and Ceylan [20]	2012	Signalization optimization	Harmony search algorithm	Simulation
D'Acierno et al. [35]	2012	Signalization optimization	ACO-based algorithm	Simulation
Kang et al. [61, 62]	2012	Highway alignment optimization	Genetic algorithm	Real world case study
Putha et al. [91]	2012	Traffic signal optimization	Ant colony, GA	Numerical example
Balseiro et al. [8]	2011	Vehicle routing	Ant colony	Numerical example
Geroliminis et al. [44]	2011	Transit mobile repair units allocation	Genetic algorithm	Real world case study
Mesbah et al. [82]	2011	Transit priority	Genetic algorithm	Numerical example
Deshpande et al. [122]	2010	Scheduling pavement rehabilitation	Multi-objective genetic algorithm	Numerical example
García-Nietoa et al.	2010	Traffic light scheduling	PSO algorithm	Simulation
Kepaptsoglou et al. [68]	2010	Pricing policy optimization	Genetic algorithm	Real world case study
Meng and Khoo [81]	2010	Ramp metering	NSGA-II	Real world case study

(continued)

Table 1 (continued)

Authors	Date	Problem	CI method	Data
Pishvaee et al. [90]	2010	Logistics network design	Memetic algorithm	Numerical example
Shimamoto et al. [96]	2010	Transit network design	NSGA-II	Ticket-based travel data[a]
Kang et al. [63]	2009	Highway alignment optimization	Genetic algorithm	Real world case study
Karlaftis et al. [65]	2009	Vehicle routing	Genetic algorithm	Real world case study
Kepaptsoglou and Karlaftis [67]	2009	Transit network design	Genetic algorithm	Real world case study
Lau et al. [72]	2009	Vehicle routing	Genetic algorithm, fuzzy algorithm	Simulation

[a]Big data applications

Another domain of CI that has attracted significant attention in transportation and traffic engineering is agent based modeling. Agent and multi-agent systems have been applied to many traffic and transportation fields including dynamic routing and congestion management. Chen et al. [24] and Bazzan and Klüge [12] reviewed the literature related to agent-based traffic modelling and simulation, and agent-based traffic control and management. However, as stated in Bazzan [11], the "agentification" of transportation problems may hinder several challenging issues (e.g. the number of agents is high, the extent and magnitude of collective behavioral patterns is immense and probably unpredictable etc.) that should be carefully examined and taken into consideration.

A significant portion of literature refers to the optimization of leaning processes involved in transportation models. Learning from extensive transportation and traffic datasets involve multi-source data distributed in many different locations and involve too many data points and extensive spatial coverage. Learning strategies inside traffic and transportation prediction models, as well as dimensionality reduction approaches and imputation problems have been systematically addressed using computationally intelligent techniques [23, 52, 80, 105, 107, 108, 110, 112, 118].

The analysis of literature indicates that there are very few big data applications to transportation optimization problems that are treated with CI methods. Shimamoto et al. [96] introduce a NSGA II algorithm to solve the transit assignment problem using ticket-based travel data. Levin and Kanza [73] implemented heuristic algorithms for the vehicle outing problem using location based data. Danalet et al. [36] leveraged campus wi-fi data to solve the pedestrian routing problem, whereas Chen et al. [26] used GPS traces and location based data for trip planning. The limited number of studies on transportation optimization using big data does not signify limited interest on the specific subject, but reflects two distinct challenges: first, large-scale optimization problems involving a significant number of modeling parameters are difficult to be estimated in a global search context; even CI that are more robust that classical approaches, may fail or become extremely time consuming, especially in a multi-objective framework [128]. Second, transportation optimization problems are complex and involve a tedious procedure for evaluating the quality of solutions when dealing with global population based search algorithms.

6 Opportunities and Challenges

6.1 The Changing Nature of Transportation Problems

Conceptually, the methodological change that big data brings to transportation is the need to automatically process and analyze data. This has significant effects on the knowledge that may be or needs to be extracted from the available data. Several solutions to problems in transportation science that were founded on static univariate data may not be applicable to dynamically changing multivariate datasets leading to the need to reexamine several phenomena or even change the way we think of transportation problems.

Three promising research fields that will most likely benefit from the data deluge area are:

- User experience mining for improving transportation services,
- Naturalistic driving experiments for monitoring driver's behavior, constructing driver's profile and identifying risk in driving, and
- Autonomous driving for congestion mitigation and safety.

The deluge of big data may not signify that some scientific questions are to be better modeled, but, a more detailed modeling approach to various phenomena may be accomplished [97]; some examples are OD surveys home interviews, census surveys, and so on. The ability to monitor the transportation and traffic related characteristics of individual road users will significantly affect the manner transportation research problems are articulated. Nevertheless, to turn data into knowledge some old dilemmas and challenges extend to big data science. These refer to model selection, real time operation, the quality and availability of the data, the quality of optimization solutions, the inference mechanisms, as well as ethical and social issues.

6.2 Big Data Analytics Versus Models

The changing nature of transportation problems often drives the need to test and evaluate new modeling paradigms robust to big data and imperfections. CI and data mining has taken a large part of the related transportation literature frequently leaving less ground to classical statistics and models. This may hinder the danger to consider that models, either statistics or borrowed by laws of physics, traditionally used to treat transportation problems are now obsolete. The truth is rather in the middle and relates to the type and extent of information needed. Evidently, a deeper understanding of the transportation problems will dictate the use of models that may translate data into causal relationships. Towards this direction, literature has emphasized the need to develop synergies with statistics to enhance the explanatory power of many CI applications [66]. Statistics may enhance the inference mechanisms of CI approaches and assure the reliability of the models developed and their generalization power.

6.3 From Batch to Real-Time Computations

The challenging task in big data analysis is not only to produce knowledge, but to produce it in a timely manner. The time to produce results relates to the size and the complexity of the datasets. Processes that may take long, but can claim increase accuracy and reliability are of limited use, if they are provided with delay. Batch model building with either data mining or statistical approaches has been the dominant approach to transportation problems. Modeling has been traditionally based on historical data, that where leveraged using different modeling paradigms to extract knowledge. In this framework, by the time new data arrive, these were batch processed to produce the output. This approach seems to be conceptually at arms with the computational needs of modern ITS systems that require timely and accurate information to disseminate to centers and users in a highly dynamic transportation environment. Data driven ITS and individual driven ITS systems are founded on real time computations, developing real-time new models that may not only respond in real-time, but learn to change their behavior in real-time (retraining strategies for CI short-term forecasting models) [108, 117]. In such conditions, optimization challenges are numerous and involve optimizing models to include new phenomena and forget past—probably incorrect or trivial—knowledge.

6.4 Data Quality, Availability, Representativeness and Relevance

Data unceasingly coming from multiple sources, at a variety of forms and in high resolutions are inhomogeneous and may contain noise and erroneous values. Noise and errors mask the significant information hindering in the data. The usual approach is to filter and apply data reduction techniques to eliminate the effect of noise and errors [48]. Data cleaning is a long standing problem but with significance in cases of big data. Data cleaning may include several tasks, such as irregularities (anomalies) detection, incomplete data filling, duplicates removal, conflicting values detection and so on. Nevertheless, these tasks are not so easy to be accomplished in the big data framework [121]; first, because many data cleaning strategies are not suitable for big data, and second, because in the big data framework, many error types (incomplete data, missing data, erroneous data duplicate data etc.) coexist, while existing techniques are focused on treating a specific error type at a time.

Furthermore, there is a thin line between information and extreme data. Noise and extreme values may contain useful information for the phenomenon under investigation. The use of advanced techniques to automatically preprocess the data and transform them to a more "analyzable" form may lead to datasets that have significantly distorted information about real world conditions [109].

Having large datasets may not always mean having a representative sample to study a phenomenon. Quality is linked to the sample size that needs to be accounted for. The collected data may account for a small part of the phenomenon both spatially and demographically. A typical example is data gathered from tweets and Facebook posts; those that do not possess a profile in social media will not be captured and included in big datasets.

The big data frequently dictate the modeling approach to follow. Nevertheless caution should be given to the modeling strategy; the belief that analyses suited for small datasets may be done with the same or better accuracy to larger datasets is misleading. There are models that have traditionally work well for small datasets, but could become unfeasible with more massive data, whereas in some modeling cases with clear underlying dynamics, simple models, such as linear regression with distinct causal implications could approximate with comparable accuracy and effectiveness the given data. Hand [48] defines the unintelligent data analysis as the one that results to over-specified models or over-idealized problems and underlines that intelligent analysis is dependent of a "good" strategy that defines the steps, decisions and actions taken to analyze a given dataset.

6.5 Inference from Data: Correlations and Causation

In the era of "big data" several researchers may claim that correlations will be enough to provide information and a deeper look to causations that may help researcher to acquire a thorough understanding of the different phenomena may not be necessary. This misconception deriving from data enthusiasts is tricky and contradicts the true intentions of data analytics. With data analytics we aim to extract information for making better and more informed decisions. Such decision based solely on correlations and deprived from causalities may be far from being accurate and intelligent.

Even if CI approaches are to be implemented, interpretation remain a focal point in transportation engineering. CI using big data can easily reveal correlations; the larger the datasets the greater number of correlations between different variables may be revealed. This does not, however, imply that causations may be achieved [116]. Moreover, several correlations may be also coincidental (spurious) [1]. The lack of straightforward inference mechanisms in CI approaches may lead to misinterpretations and erroneous results. This is a major shortcoming of applying CI methods to transportation data and should be taken into consideration. Big data are complex and causation can be distorted by various factors such as latent variables, indirect influences imposed by various systems acting simultaneously, multi-collinearity, missing values and so on.

6.6 Quality of Optimization Solutions and Uncertainties

Evaluating optimization solutions is a time consuming and costly task. The more complex the optimization problems the less efficient the global population based approaches become. To reduce the time and effort needed to provide optimization solutions of high quality, surrogate modeling often qualifies as a viable solution. Surrogate modeling is a macro-modeling technique that aims to minimize the time and computational load to develop simulations to replicate input-output relationships [41]. The aim is to produce a faster and simpler approximation of a simulator to make optimization, design space exploration, etc. feasible.

Another critical issue to consider is the robustness of the produced solutions over time. Most transportation phenomena has significant spatio-temporal dependencies that may influence the quality and consistency of the produced solutions. As such, robustness over time is a critical characteristics of the optimization strategies. This may be tackled by selecting the optimization approach that produces results that are the least affected by the varying conditions (changes in variables etc.). The use of dynamic optimization strategies that are computationally intensive seem to be out of the context of real-time ITS applications. Evidently, achieving a tradeoff between the best solution and the optimum solution over time—that will change only when a solution will provide results that are no longer acceptable—is a viable approach [57, 128].

6.7 Ethics, Privacy, Inequalities

The big data deluge in transportation comes with significant ethical and institutional challenges. As in all disciplines, big data, especially those coming from participatory sensing, have serious ethical and privacy issues that are frequently addressed but rarely understood. Nowadays, a legislative framework that will dictate the ethical boundaries of using personal data streams is missing.

Moreover, until recently, data was a key advantage of a scientific work because several phenomena, especially those dealing with behavioral aspects, were difficult to be monitored. Nowadays, having data still provides a competitive advantage, but for different reasons. Although the technological means to achieve a detailed monitoring of complex phenomena exist, they are not accessible to everyone. The digital divides created by those who possess technology and data are significant for achieving innovation [14]. Moreover, inequalities will progressively extend to research institutions and Academia between those that may fund big data systems and those that do not possess the economic means to penetrate the market of big data and use them to their benefit. Significant competitive advantage will have those companies and organizations that may not only possess big data, but also can analyze them.

7 The Road Ahead

In the near future, every moving object (both humans and machines) is planned to have a unique identity and operate in a smart social and environmental setting. In this framework, advanced skills in data analytics and optimization will be required to solve complex problems and materialize advanced transportation ideas. The road ahead contains CI, but they have to be applied with caution. Some drivers for success will be: i. develop real-time modeling efforts and efficient solutions to complex phenomena and settings, ii. the development of synergies and the use of intuition to enhance explanatory power, iii. the development of test beds and test data to battle inequalities and evaluate ongoing development, iv. the integration of nature inspired algorithms to enable the full abilities of CI, v. cloud and parallel computing for increasing computational power and reducing the cost of transportation services, and vi. the development of new educational paradigms so as transportation researchers and practitioners can cope with the demanding algorithms for treating big data.

The rapid growing of transportation data impose delivering computations and results that reflect the dynamically evolving transportation phenomena in real-time. In this framework researchers should focus on responsive new methods and model building techniques. Moreover, the spatio-temporal complexity seen in most transportation datasets impose the decomposition of a problem to many simpler ones; this decomposition should extend to model building. Ensembles of models rather than a single approach should be evaluated to deliver reliable and accurate models and predictions. As for optimization, literature review underlined that although CI global optimization techniques may well cope with the complexities seen in transportation datasets, they have been rarely used in big transportation data due to the high computational cost they entail. It is of great importance to use big data to develop more flexible and computationally less costly CI meta-optimization techniques—for example surrogates—or improve the manner to formulate optimization problems.

The rise of CI techniques to handle big data does not make statistics obsolete. Several researchers have systematically underlined that the statistical thinking is the means to justify the inferential leap from data to knowledge. Possible synergies between these two different schools of thought will increase the explanatory power of CI models and their transparency [66]. Statistics may be useful for enhancing the clarity about the modeling goals, assessing for the reliability of the model developed, accounting for sources of uncertainty in the underlying data mechanism and models [45]. In the model development and evaluation stages, statistics can provide the theoretical means for testing for optimality and suitability of the learning algorithms. Moreover, statistics may be used to extract causalities, if necessary, an issue largely disregarded in the CI literature. In this spirit, intuition has a great role to play in the understanding of the huge streams of data. The CI approaches should be tied to human intuition so as results to be reflect reality and not a myopic look at the different phenomena.

Research using big data in transportation should be supported by publically available testbeds and test data. Test beds of varying size and complexity are a critical tool for reducing inequalities, supporting innovation, but also evaluating ongoing research and may serve as a proof-of-concept tool [115]. To this end, open data is considered today as the greatest enabler of research in intelligent transportation systems. A typical example of the direction towards freely available data is the European Open Government Data Initiative (EU OGDI). This initiative targets to create a transparent environment without discrimination and exclusivity constraints where both data and software can be freely stored to improve practices and implemented policies across EU member countries. The concept of open big data multiplies the sources of creativity and collective innovation, as new applications and algorithms are produced by both established providers (e.g. Google, IBM, SAS etc.) and public authorities, but also by individual initiatives from programmers (e.g. applications on smart phones).

Another critical issue that will dictate the future of CI in transportation is the ability to fully benefit from artificial intelligence (AI), a key technology to improve the efficiency, safety, and environmental-compatibility of transportation systems [92]. Until now, CI and AI applications have been limited to specific modules of ITS applications, especially for data analysis and prediction disregarding their powerful capabilities for data managing and decision making [32]. Extended usage of CI and AI techniques is needed to fully benefit from their unique capabilities. Towards this direction, concepts such as cloud (computation, software, data access, and storage services) that do not require end-user knowledge of the physical location and configuration of the system that delivers the services, and parallel computing (clusters of computers), can enable the implementation of complex network level ITS [50, 120, 126].

In the instrumented future, transportation engineers and researchers are challenged to be capable of applying both transportation science and interdisciplinary data analyses for the realization and evaluation of their advanced ideas. Evidently, the advent of the new "big data" area in transportation dictates the need to develop new educational paradigms to produce qualified transportation researchers and practitioners able cope with the demanding algorithms for treating big data. The aim is not to replace other disciplines but to be able to produce engineers that may understand and efficient use the full potential of big datasets and the accompanying modeling tools.

Acknowledgments This work is part of research co-financed by the European Union (European Social Fund—ESF) and the Hellenic National Funds, through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)—Research Funding Program "Aristeia I". This paper is dedicated to the memory of my mentor and friend, Professor Matthew G. Karlaftis.

References

1. Aldrich J (1995) Correlations genuine and spurious in Pearson and Yule. Stat Sci 10:364–376
2. Arango C, Cortés P, Onieva L, Escudero A (2013) Simulation-optimization models for the dynamic berth allocation problem. Comput Aided Civil Infrastruct Eng 28(10):769–779
3. Astarita V, Bertini L, d'Elia S, Guido G (2006) Motorway traffic parameter estimation from mobile phone counts. Eur J Oper Res 175:1435–1446
4. Atzori L, Iera A, Morabito F (2010) The internet of things: a survey. Comput Netw 54(15):2787–2805
5. Aupetit S, Riff J, Buttelli O, Espié S (2013) Naturalistic study of rider's behaviour in initial training in France: evidence of limitations in the educational content. Accid Anal Prev 58:206–217
6. Ayuso M, Guillén M, Pérez-Marín AM (2014) Time and distance to first accident and driving patterns of young drivers with pay-as-you-drive insurance. Accid Anal Prev 73:125–131
7. Bai Q, Ahmed A, Li Z, Labi S (2014) A hybrid pareto frontier generation method for trade-off analysis in transportation asset management. Comput Aided Civil Infrastruct Eng
8. Balseiro SR, Loiseau I, Ramonet J (2011) An ant colony algorithm hybridized with insertion heuristics for the time dependent vehicle routing problem with time windows. Comput Oper Res 38(6):954–966
9. Bar-Gera H (2007) Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel. Transp Res Part C 15:380–391
10. Basyoni Y, Talaat H (2014) A bi-level traffic data extraction procedure via cellular phone network for inter-city travel. J Intell Transp Syst Technol Plann Oper (forthcoming)
11. Bazzan AL (2009) Opportunities for multiagent systems and multiagent reinforcement learning in traffic control. Auton Agent Multi-Agent Syst 18(3):342–375
12. Bazzan AL, Klügl F (2014) A review on agent-based technology for traffic and transportation. Knowl Eng Rev 29(03):375–403
13. Bierlaire M, Chen J, Newman J (2013) A probabilistic map matching method for smartphone GPS data. Transp Res Part C: Emerg Technol 26:78–98
14. Boyd D, Crawford K (2012) Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. Inf Commun Soc 15(5):662–679
15. Breiman L (2001) Statistical modeling: the two cultures (with comments and a rejoinder by the author). Stat Sci 16(3):199–231
16. Caceres N, Wideberg JP, Benitez FG (2007) Deriving origin-destination data from a mobile phone network. Intell Transp Syst 1(1):15–26
17. Cai H, Jia X, Chiu AS, Hu X, Xu M (2014) Siting public electric vehicle charging stations in Beijing using big-data informed travel patterns of the taxi fleet. Transp Res Part D: Transp Environ 33:39–46
18. Caragliu A, Del Bo C, Nijkamp P (2009) Smart cities in Europe. Serie Research Memoranda 0048 (VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics)
19. Castro PS, Zhang D, Li S (2012) Urban traffic modelling and prediction using large scale taxi GPS traces. In: Pervasive computing. Springer, Berlin, Heidelberg, pp 57–72
20. Ceylan H, Ceylan H (2012) A hybrid harmony search and TRANSYT hill climbing algorithm for signalized stochastic equilibrium transportation networks. Transp Res Part C: Emerg Technol 25:152–167
21. Ceylan H (2006) Developing combined genetic algorithm hill-climbing optimization method for area traffic control. J Transp Eng ASCE 132(8):663–671
22. Ceylan H, Bell MGH (2004) Traffic signal timing optimization based on genetic algorithm approach, including drivers' routing. Transp Res Part B 38(4):329–342
23. Chan KY, Dillon T, Chang E, Singh J (2013) Prediction of short-term traffic variables using intelligent swarm-based neural networks. IEEE Trans Control Syst Technol 21(1):263–274
24. Chen B, Cheng HH (2010) A review of the applications of agent technology in traffic and transportation systems. IEEE Trans Intell Transp Syst 11(2):485–497

25. Chen C, Zhang D, Guo B, Ma X, Pan G, Wu Z (forthcoming) TripPlanner: personalized trip planning leveraging heterogeneous crowdsourced digital footprints. IEEE Trans Intell Transp Syst
26. Chen C, Zhang D, Li N, Zhou ZH (2014) B-Planner: planning bidirectional night bus routes using large-scale taxi GPS traces. IEEE Trans Intell Transp Syst 15(4):1451–1465
27. Cheng Z, Caverlee J, Lee K (2010) You are where you tweet: a content-based approach to geo-locating twitter users. In: Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, Toronto, ON, Canada
28. Cheng Z, Caverlee J, Lee K, Sui D (2011) Exploring millions of footprints in location sharing services. In: Proceedings of the fifth international conference on weblogs and social media. AAAI, Barcelona, Spain
29. Cheng S, Shi Y, Qin Q, Bai R (2013) Swarm intelligence in big data analytics. In: Intelligent data engineering and automated learning-IDEAL 2013. Springer, Berlin, Heidelberg, pp 417–426
30. Chevrier R, Pellegrini P, Rodriguez J (2013) Energy saving in railway timetabling: a bi-objective evolutionary approach for computing alternative running times. Transp Res Part C: Emerg Technol 37:20–41
31. Chira C, Sedano J, Villar JR, Cámara M, Corchado E (2014) Urban bicycles renting systems: modelling and optimization using nature-inspired search methods. Neurocomputing 135:98–106
32. Chowdhury M, Sadek AW (2012) Advantages and limitations of artificial intelligence. Artif Intell Appl Crit Transp Issues 6
33. Collins C, Hasan S, Ukkusuri SV (2013) A novel transit rider satisfaction metric. J Public Transp 16(2):21–45
34. Cong Z, De Schutter B, Babuška R (2013) Ant colony routing algorithm for freeway networks. Transp Res Part C: Emerg Technol 37:1–19
35. D'Acierno L, Gallo M, Montella B (2012) An ant colony optimisation algorithm for solving the asymmetric traffic assignment problem. Eur J Oper Res 217(2):459–469
36. Danalet A, Farooq B, Bierlaire M (2014) A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures. Transp Res Part C: Emerg Technol 44:146–170
37. Doolan R, Muntean GM (2014) Time-ants: an innovative temporal and spatial ant-based vehicular routing mechanism. In: Intelligent vehicles symposium proceedings, 2014 IEEE. IEEE, pp 951–956
38. Fagnant DJ, Kockelman KM (2014) The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. Transp Res Part C: Emerg Technol 40:1–13
39. Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. AI Mag 17(3):37
40. Forcael E, González V, Orozco F, Vargas S, Pantoja A, Moscoso P (2014) Ant colony optimization model for tsunamis evacuation routes. Comput Aided Civil Infrastruct Eng 29(10):723–737
41. Forrester A, Sobester A, Keane A (2008) Engineering design via surrogate modelling: a practical guide. Wiley, Chichester
42. Friedrich M, Immisch K, Jehlicka P, Otterstatter T, Schlaich J (2010) Generating origin-destination matrices from mobile phone trajectories. Transp Res Rec 2196:93–101
43. Galland S, Knapen L, Yasar AUH, Gaud N, Janssens D, Lamotte O, Wets G (2014) Multi-agent simulation of individual mobility behavior in carpooling. Transp Res Part C: Emerg Technol
44. Geroliminis N, Kepaptsoglou K, Karlaftis MG (2011) A hybrid hypercube-genetic algorithm approach for deploying many emergency response mobile units in an urban network. Eur J Oper Res 210(2):287–300
45. Glymour C, Madigan D, Pregibon D, Smyth P (1997) Statistical themes and lessons for data mining. Data Min Knowl Discov 1(1):11–28

46. Goksal FP, Karaoglan I, Altiparmak F (2013) A hybrid discrete particle swarm optimization for vehicle routing problem with simultaneous pickup and delivery. Comput Ind Eng 65(1):39–53
47. Guido G, Gallelli V, Saccomanno F, Vitale A, Rogano D, Festa D (2014) Treating uncertainty in the estimation of speed from smartphone traffic probes. Transp Res Part C: Emerg Technol 47:100–112
48. Hand DJ (2000) Data mining. New challenges for statisticians. Soc Sci Comput Rev 18(4):442–449
49. Hasan S, Ukkusuri SV (2014) Urban activity pattern classification using topic models from online geo-location data. Transp Res Part C: Emerg Technol 44:363–381
50. Hashem IAT, Yaqoob I, Anuar NB, Mokhtar S, Gani A, Khan SU (2015) The rise of "big data" on cloud computing: review and open research issues. Inf Syst 47:98–115
51. Herrera JC, Work DB, Herring R, Ban X, Jacobson Q, Bayen A (2010) Evaluation of traffic data obtained via GPS-enabled mobile phones: the mobile century field experiment. Transp Res Part C 18:568–583
52. Hong WC (2012) Application of seasonal SVR with chaotic immune algorithm in traffic flow forecasting. Neural Comput Appl 21(3):583–593
53. Hood J, Sall E, Charlton B (2011) A GPS-based bicycle route choice model for San Francisco. California. Transp Lett 3(1):63–75
54. Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. Neural Netw 2(5):359–366
55. Jha MK, Schonfeld P (2004) A highway alignment optimization model using geographic information systems. Transp Res Part A: Policy Pract 38(6):455–481
56. Jia L, Feng X, Zou G (2013) Solving multiobjective bilevel transportation-distribution planning problem by modified NSGA II. In: 2013 9th international conference on computational intelligence and security (CIS). IEEE, pp 303–307
57. Jin Y, Tang K, Yu X, Sendhoff B, Yao X (2013) A framework for finding robust optimal solutions over time. Memetic Comput 5(1):3–18
58. Jin Y, Sendhoff B (2009) A systems approach to evolutionary multiobjective structural optimization and beyond. IEEE Comput Intell Mag 4(3):62–76
59. Kallioras NA, Lagaros ND, Karlaftis MG (2013) An improved harmony search algorithm for emergency inspection scheduling. Eng Optim (ahead-of-print) 1–23
60. Kammoun HM, Kallel I, Casillas J, Abraham A, Alimi AM (2014) Adapt-Traf: an adaptive multiagent road traffic management system based on hybrid ant-hierarchical fuzzy model. Transp Res Part C: Emerg Technol 42:147–167
61. Kang C, Ma X, Tong D, Liu Y (2012) Intra-urban human mobility patterns: an urban morphology perspective. Phys A: Stat Mech Appl 391(4):1702–1717
62. Kang MW, Jha MK, Schonfeld P (2012) Applicability of highway alignment optimization models. Transp Res Part C: Emerg Technol 21(1):257–286
63. Kang MW, Schonfeld P, Yang N (2009) Prescreening and repairing in a genetic algorithm for highway alignment optimization. Comput Aided Civil Infrastruct Eng 24(2):109–119
64. Karlaftis MG, Vlahogianni EI (2009) Memory properties and fractional integration in transportation time-series. Transp Res Part C: Emerg Technol 17(4):444–453
65. Karlaftis MG, Kepaptsoglou K, Sambracos E (2009) Containership routing with time deadlines and simultaneous deliveries and pick-ups. Transp Res Part E: Logist Transp Rev 45(1):210–221
66. Karlaftis MG, Vlahogianni EI (2011) Statistics versus neural networks in transportation research: differences, similarities and some insights. Transp Res Part C: Emerg Technol 19(3):387–399
67. Kepaptsoglou K, Karlaftis MG (2009) The bus bridging problem in metro operations: conceptual framework, models and algorithms. Public Transp 1(4):275–297
68. Kepaptsoglou K, Karlaftis MG, Li Z (2010) Optimizing pricing policies in Park-and-Ride facilities: a model and decision support system with application. J Transp Syst Eng Inf Technol 10(5):53–65

69. Kontou E, Kepaptsoglou K, Charalampakis AE, Karlaftis MG (2014) The bus to depot allocation problem revisited: a genetic algorithm. Public Transp 6(3):237–255
70. Lagaros ND, Kepaptsoglou K, Karlaftis MG (2012) Fund allocation for civil infrastructure security upgrade. J Manage Eng 29(2):172–182
71. Laney D (2001) 3D data management: controlling data volume, velocity and variety. Gartner. Accessed 6 Feb
72. Lau HC, Chan TM, Tsui WT, Chan FT, Ho GT, Choy KL (2009) A fuzzy guided multi-objective evolutionary algorithm model for solving transportation problem. Expert Syst Appl 36(4):8255–8268
73. Levin R, Kanza Y (2014) TARS: traffic-aware route search. GeoInformatica 18(3):461–500
74. Liang X, Zheng X, Lv W, Zhu T, Xu K (2012) The scaling of human mobility by taxis is exponential. Phys A: Stat Mech Appl 391(5):2135–2144
75. Lin DY, Ku YH (2014) Using genetic algorithms to optimize stopping patterns for passenger rail transportation. Comput Aided Civil Infrastruct Eng 29(4):264–278
76. Liu Y, Kang C, Gao S, Xiao Y, Tian Y (2012) Understanding intra-urban trip patterns from taxi trajectory data. J Geogr Syst 14(4):463–483
77. Liu Y, Li P, Wehner K, Yu J (2013) A generalized integrated corridor diversion control model for freeway incident management. Comput Aided Civil Infrastruct Eng 28(8):604–620
78. Liu Y, Roshandeh AM, Li Z, Kepaptsoglou K, Patel H, Lu X (2014) Heuristic approach for optimizing emergency medical services in road safety within large urban networks. J Transp Eng 140(9):04014043
79. Liu Y, Wang F, Xiao Y, Gao S (2012) Urban land uses and traffic 'source-sink areas': evidence from GPS-enabled taxi data in Shanghai. Landscape Urban Plann 106(1):73–87
80. Lv Y, Duan Y, Kang W, Li Z, Wang FY (2014) Traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transp Syst (forthcoming)
81. Meng Q, Khoo HL (2010) A Pareto-optimization approach for a fair ramp metering. Transp Res Part C: Emerg Technol 18(4):489–506
82. Mesbah M, Sarvi M, Currie G (2011) Optimization of transit priority in the transportation network using a genetic algorithm. IEEE Trans Intell Transp Syst 12(3):908–919
83. Musicant O, Bar-Gera H, Schechtman E (2010) Electronic records of undesirable driving events. Transp Res Part F: Traffic Psychol Behav 13(2):71–79
84. Musicant O, Bar-Gera H, Schechtman E (2014) Temporal perspective on individual driver behavior using electronic records of undesirable events. Accid Anal Prev 70:55–64
85. Paefgen J, Staake T, Fleisch E (2014) Multivariate exposure modeling of accident risk: insights from Pay-as-you-drive insurance data. Transp Res Part A: Policy Pract 61:27–40
86. Paefgen J, Staake T, Thiesse F (2013) Evaluation and aggregation of pay-as-you-drive insurance rate factors: a classification analysis approach. Decis Support Syst 56:192–201
87. Pahlavani P, Delavar MR (2014) Multi-criteria route planning based on a driver's preferences in multi-criteria route selection. Transp Res Part C: Emerg Technol 40:14–35
88. Pan CX, Lu JG, Di S, Ran B (2006) Cellular-based data-extracting method for trip distribution. Traffic Urban Data 1945:33–39
89. Papinski D, Scott DM, Doherty ST (2009) Exploring the route choice decision-making process: a comparison of planned and observed routes obtained using person-based GPS. Transp Res Part F: Traffic Psychol Behav 12(4):347–358
90. Pishvaee MS, Farahani RZ, Dullaert W (2010) A memetic algorithm for bi-objective integrated forward/reverse logistics network design. Comput Oper Res 37(6):1100–1112
91. Putha R, Quadrifoglio L, Zechman E (2012) Comparing ant colony optimization and genetic algorithm approaches for solving traffic signal coordination under oversaturation conditions. Computer Aided Civil Infrastruct Eng 27(1):14–28
92. Sadek AW (ed) (2007) Artificial intelligence in transportation: information for application. Transportation Research Board Circular (E-C113), TRB, National Research Council, Washington, DC. http://onlinepubs.trb.org/onlinepubs/circulars/ec113.pdf
93. Sarle WS (1994) Neural networks and statistical models. In: Proceedings of the nineteenth annual SAS users group international conference, 1–13 April

94. Shafahi Y, Bagherian M (2013) A customized particle swarm method to solve highway alignment optimization problem. Computer Aided Civil Infrastruct Eng 28(1):52–67
95. Shichrur R, Sarid A, Ratzon NZ (2014) Determining the sampling time frame for in-vehicle data recorder measurement in assessing drivers. Transp Res Part C: Emerg Technol 42:99–106
96. Shimamoto H, Murayama N, Fujiwara A, Zhang J (2010) Evaluation of an existing bus network using a transit network optimisation model: a case study of the Hiroshima city bus network. Transportation 37(5):801–823
97. Shuldiner AT, Shuldiner PW (2013) The measure of all things: reflections on changing conceptions of the individual in travel demand modeling. Transportation 40(6):1117–1131
98. Stolfi DH, Alba E (2014) Red swarm: reducing travel times in smart cities by using bio-inspired algorithms. Appl Soft Comput 24:181–195
99. Sun D, Benekohal RF, Waller ST (2006) Bi-level programming formulation and heuristic solution approach for dynamic traffic signal optimization. Comput Aided Civil Infrastruct Eng 21(5):321–333
100. Takeda K, Miyajima C, Suzuki T, Angkititrakul P, Kurumida K, Kuroyanagi Y, Komada Y (2012) Self-coaching system based on recorded driving data: learning from one's experiences. IEEE Trans Intell Transp Syst 13(4):1821–1831
101. Teklu F, Sumalee A, Watling D (2007) A genetic algorithm approach for optimizing traffic control signals considering routing. Comput Aided Civil Infrastruct Eng 22(1):31–43
102. Teodorović D (2008) Swarm intelligence systems for transportation engineering: principles and applications. Transp Res Part C: Emerg Technol 16(6):651–667
103. Terzi S, Serin S (2014) Planning maintenance works on pavements through ant colony optimization. Neural Comput Appl 25:1–11
104. Toledo T, Musicant O, Lotan T (2008) In-vehicle data recorders for monitoring and feedback on drivers' behavior. Transp Res Part C: Emerg Technol 16(3):320–331
105. Tselentis DI, Vlahogianni EI, Karlaftis MG (2014) Improving short-term traffic forecasts: to combine models or not to combine? IET Intell Transp Syst (forthcoming)
106. Vlacheas P, Giaffreda R, Stavroulaki V, Kelaidonis D, Foteinos V, Poulios G, Demestichas P, Somov A, Biswas AR, Moessner K (2013) Enabling smart cities through a cognitive management framework for the internet of things. IEEE Commun Mag 51(6):102–111
107. Vlahogianni EI, Karlaftis MG (2012) Comparing freeway lane speed patterns under fine and adverse weather conditions. Nonlinear Dyn 69(4):1949–1963
108. Vlahogianni EI (2009) Enhancing predictions in signalized arterials with information on short-term traffic flow dynamics. J Intell Transp Syst: Technol Plann Oper 13(2):73–84
109. Vlahogianni EI, Karlaftis MG (2011) Aggregating temporal and spatial data: implication for statistical characteristics and model choice. Transp Lett: Int J Transp Res 3(1):37–49
110. Vlahogianni EI, Karlaftis MG (2013) Testing and comparing neural network and statistical approaches for predicting transportation time series. Transp Res Rec: J Transp Res Board 2399(1):9–22
111. Vlahogianni EI, Golias JC, Karlaftis MG (2004) Short-term traffic forecasting: overview of objectives and methods. Transp Rev 24(5):533–557
112. Vlahogianni EI, Karlaftis MG, Golias JC (2005) Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. Transp Res Part C: Emerg Technol 13(3):211–234
113. Vlahogianni EI, Karlaftis MG, Golias JC (2006) Statistical methods for detecting nonlinearity and non-stationarity in univariate short-term time-series of traffic volume. Transp Res Part C: Emerg Technol 14(5):351–367
114. Vlahogianni EI, Karlaftis MG, Golias JC (2008) Temporal evolution of short-term urban traffic flow: a nonlinear dynamics approach. Comput Aided Civil Infrastruct Eng 23(7):536–548
115. Vlahogianni EI, Karlaftis MG, Golias JC (2014) Short-term traffic forecasting: where we are and where we're going. Transp Res Part C: Emerg Technol 43:3–19
116. Vlahogianni EI, Karlaftis MG, Orfanou FP (2012) Modeling the effects of weather and traffic on the risk of secondary incidents. J Intell Transp Syst 16(3):109–117

117. Vlahogianni EI, Karlaftis MG, Golias JC, Kourbelis ND (2006) Pattern-based short-term urban traffic predictor. In: Intelligent transportation systems conference, 2006, ITSC'06. IEEE, pp 389–393
118. Vlahogianni EI, Yannis G, Golias JC (2013) Critical power two wheeler riding patterns at the emergence of an incident. Accid Anal Prev 58:340–345
119. Vlahogianni EI, Yannis G, Golias JC (2014) Detecting powered-two-wheeler incidents from high resolution naturalistic data. Transp Res Part F: Traffic Psychol Behav 22:86–95
120. Wang FY (2010) Parallel control and management for intelligent transportation systems: concepts, architectures, and applications. IEEE Trans Intell Transp Syst 11(3):630–638
121. Wang H, Li M, Bu Y, Li J, Gao H, Zhang J (2014) Cleanix: a big data cleaning parfait. In: Proceedings of the 23rd ACM international conference on conference on information and knowledge management. ACM, pp 2024–2026
122. Yang F, Jin PJ, Cheng Y, Ran B (2014) Origin-destination estimation for non-commuting trips using location-based social networking data. Int J Sustain Transp (just-accepted)
123. Yang N, Kang MW, Schonfeld P, Jha MK (2014) Multi-objective highway alignment optimization incorporating preference information. Transp Res Part C: Emerg Technol 40:36–48
124. Yin W, Murray-Tuite P, Ukkusuri SV, Gladwin H (2014) An agent-based modeling system for travel demand simulation for hurricane evacuation. Transp Res Part C: Emerg Technol 42:44–59
125. Zhang G, Zhang H, Li L, Dai C (2014) Agent-based simulation and optimization of urban transit system. IEEE Trans Intell Transp Syst 15(2)
126. Zhang L, Wu C, Li Z, Guo C, Chen M, Lau F (2013) Moving big data to the cloud: an online cost-minimizing approach. IEEE J Sel Areas Commun 31(12):2710–2721
127. Zhang S, Lee CKM, Chan HK, Choy KL, Wu Z (2015) Swarm intelligence applied in green logistics: a literature review. Eng Appl Artif Intell 37:154–169
128. Zhou Z, Chawla N, Jin Y, Williams G (2014) Big data opportunities and challenges: discussions from data analytics perspectives. IEEE Comput Intell Mag 9(4):62–74