Big Data Analytics for Transportation: Problems and Prospects for its Application in China

Robert P. Biuk-Aghai, Weng Tat Kou and Simon Fong
Department of Computer and Information Science
Faculty of Science and Technology
University of Macau
Email: {robertb,mb45467,ccfong}@umac.mo

***Abstract*—Transportation, as a means for moving goods and people between different locations, is a vital element of modern society. In this paper, we discuss how big data technology infrastructure fits into the current development of China, and provide suggestions for improvement. We discuss the current situation of China's transportation system, and outline relevant big data technologies that are being used in the transportation domain. Finally we point out opportunities for improvement of China's transportation system, through standardisation, integration of big data analytics in a national framework, and point to the future of transportation in China and beyond.**

I. Introduction

Transportation, as a means for moving goods and people between different locations, is a vital element of modern society. Since the earliest days of the industrial revolution, transportation has facilitated economic development by moving materials, resources, products and people [1]. However, transportation is affected by problems such as road traffic congestion. Governments attempt to formulate effective road projects that improve the traffic situation and overcome existing traffic problems. A difficulty in formulating a good transportation strategy is the lack of empirical data of actual road conditions. Thus when transportation projects are being implemented, the government may not have planned an effective diversion policy because future traffic conditions are unknown and hard to predict. As a result, the road traffic authority faces the dilemma of having to choose between redevelopment of the whole transportation infrastructure and extending current road designs.

Transportation is also connected with other problems, including safety and environmental pollution. Traffic accidents are not only costly in terms of loss of life, but also in economic cost through damaged property (vehicles and others), emergency services, loss of income, medical costs, rehabilitation costs, legal costs etc. [2]. Environmental pollution has costs through damage to our natural environment, as well as financial costs linked to health impact, global warming and others. In China the financial cost of the health impact of air pollution was estimated to be USD 1.4 trillion in 2010, a large part of which is likely due to road traffic [3]. Since then the problem has only intensified. Governments attempt to monitor the amount of pollution emission such as CO_2 and to ensure economic growth follows pace with sustainable development.

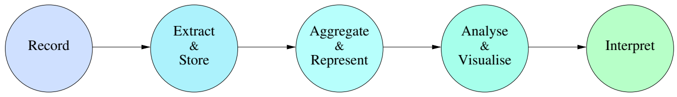

Fig. 1. Big data analytics process.

As in all other economic sectors, big data analytics is being applied in transportation. Big data is usually understood to mean data of large size, the "4 Vs": volume, velocity, variety, veracity. Big data analytics refers to the application of methods of data analysis, modelling, mining and visualisation of large datasets for sensemaking and to discover relevant and actionable patterns of interest. It typically follows a sequence of activities as depicted in Fig.1. Applying big data analytics has the potential to create more intelligent transportation systems.

Transportation in China is managed by the public administration of the government. However, big data analytics projects usually originate from universities, research centres, and their business spin-offs. The government may outsource operation and service to these businesses which deliver intelligent transportation technology. Given the Chinese government policy, such big data business is limited to domestic firms.

In this paper we discuss the problems of traffic in China and point out potential solutions. Specifically, our contributions are, firstly an analysis of the current transportion situation in China, and secondly a national multi-level transportation big data analytics framework.

II. China's Transportation System

In 1949 at the founding of the People's Republic of China, the length of railways was approximately 21,800km. In the 65 years until 2014 this increased more than 5-fold to approximately 112,000km [4], equivalent to an annual growth rate of 2.55%. China's highways and expressways have likewise witnessed a gradual expansion, improving national transportation. Table I summarises the development of China's land, water and air transportation routes over the past 10 years (source: [5]).

However, along with the economic development of China, the number of privately owned vehicles has grown dramatically in the past decade, leading increasingly to problems of traffic congestion. Fig.2 shows the unequal growth of highways

978-1-5090-0931-2/16/$31.00 ©2016 IEEE

TABLE I
LENGTH OF TRANSPORTATION ROUTES IN CHINA (UNIT: 1,000 KM)

Year	Railways in Operation	Highways	Navigable Inland Waterways	Regular Civil Aviation Routes
2005	75.4	3345.2	123.3	1998.5
2006	77.1	3457.0	123.4	2113.5
2007	78.0	3583.7	123.5	2343.0
2008	79.7	3730.2	122.8	2461.8
2009	85.5	3860.8	123.7	2345.1
2010	91.2	4008.2	124.2	2765.1
2011	93.2	4106.4	124.6	3490.6
2012	97.6	4237.5	125.0	3280.1
2013	103.1	4356.2	125.9	4106.0
2014	111.8	4463.9	126.3	4637.2

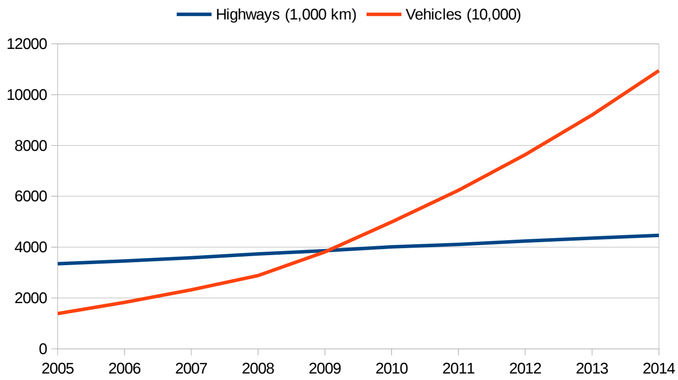

Fig. 2. Growth of length of highways (in 1,000 km) vs. number of privately owned vehicles (in 10,000).

vs. privately owned vehicles (source: [5]). It is clear that the pace of growth of highways does not keep up with that of vehicles.

Normalising the data on the length of highways and the number of privately owned vehicles with the starting year 2005, we then calculated a congestion indicator as a ratio of these two normalised values. This indicator is shown in Fig.3. The rising line indicates the rising ratio of vehicles to available road, and indicates an almost 6-fold increase in the 10-year interval 2005–2014. A trend line (linear regression) is overlaid, with $R^2 = 0.9689$, indicating a very good fit.

Thus the road transportation network in China is increasingly suffering from overcrowding, leading to decreasing satisfaction among road users. However, because of land scarcity in the more highly developed areas, there is a limit on how much new road development is possible.

To manage the traffic, intelligent urban transportation technology such as road monitoring, speed cameras, automatic toll gates etc. are being deployed to ever larger extent. According to statistics from the China Urban Intelligent Transportation market research report [6], the compound annual growth rate of urban intelligent transportation between 2008 and 2013 was 20.2%. The 12th national 5-year transportation plan predicted the 2014 annual intelligent transportation market value to reach

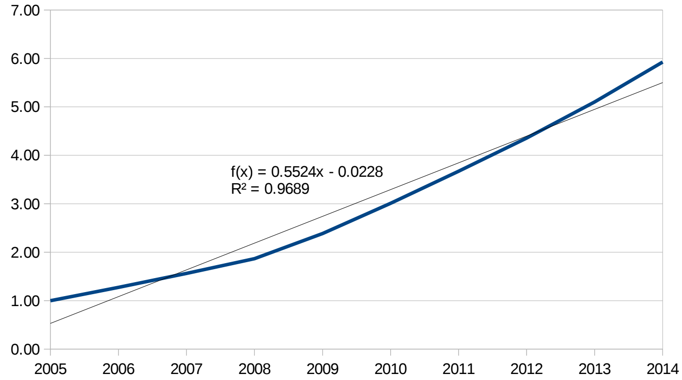

Fig. 3. Congestion indicator data (with trendline).

46 billion yuan (about USD 7 billion), and predicted the investment in intelligent transportation in the coming 10 years to amount to 170 billion yuan (about USD 26 billion). Clearly the market is huge and growing.

Although intelligent transportation technology is developing in China, there is little effort to use collected data for big data analytics. Moreover, software systems still fall short of those from other more highly developed countries. Obsolete software and inefficient infrastructure may not match international big data standards.

Currently, big data technology is considered a luxury infrastructure. Governments may not yet be ready to commit large amounts of budgets for big data investment as the return is not yet verifiable. The government should adopt a strategy of developing big data technology progressively, starting with main highways to sub-routes and city roads, which is a top-down approach. This enables the country to enjoy the benefits of big data technology as well as to develop the infrastructure that is affordable and with a reasonable cost within a given period. Given the growth of private vehicle ownership, along with the continued growth of the population, the need for developing advanced big data analytics models and systems is become increasingly pressing.

The largest problem in developing big data technology in China is related to the huge data volume captured from a very large network of traffic sensors, which requires high speed technology to capture and aggregate this data. Thus the initial investment for surveillance tools and traffic sensors is high. However, developing big data technology in China has the advantage of economies of scale. Given the size of the country, any big data technology developed for road transportation can be reused, thus ensuring a good return on investment.

The following subsections highlight some of the main problems of China's road transportation.

A. City Structure

Metropolitan centres in China are structured differently from big cities in the West: whereas in the West functional districts are often distributed around the city, in China they are often concentrated in the centre. Political, commercial and even cultural centres may all be in close proximity in the city centre.

Fig. 4. Beijing traffic jam (photo by Axel Drainville, unmodified, licensed through Creative Commons CC BY-NC 2.0, creativecommons.org).

The city of Beijing is a typical case of this, where traffic jams of 3–4 hours, or longer, occur frequently [7]. This leads to a high influx of traffic from the periphery to the centre, and consequently a concentration of heavy traffic with congestion in the city centre, such as the one shown in Fig.4.

B. Law Enforcement

Respect for traffic laws is poor among many Chinese drivers, resulting in numerous accidents with sometimes serious consequences. Regarding traffic laws, a report in The Economist found that “these laws are entirely ignored” [8]. Ignoring traffic laws was also found to be the main reason for traffic deaths in another study, which cited speeding, driving without a license, driving in the wrong lane, and drunk driving as the main causes of fatal accidents [9].

Local authorities often set up speed cameras to control the speed on their roads, and to serve as a source of revenue. However, given the size of the Chinese transportation network it is impractical to monitor all sections of all roads.

C. Sensitivity to Special Events

During special events, such as nationwide holidays (Chinese new year, Labour day holidays, October 1st “Golden Week”), massive traffic problems arise as large portions of the population travel around the country. Transportation policy makers do not sufficiently plan for contingencies and such times of peak traffic because they lack the support of empirical data. Consequently, numerous transportation problems happen. One of the worst traffic jams of all time happened near Beijing on the Beijing-Tibet expressway, stretching 100km and lasting over 10 days [10]. It was caused by roadworks on a section of the expressway, and was worsened when days later parts of another nearby road were closed. This inappropriate action by the local traffic authorities lacking an overall view of the situation led to this exaggerated scenario.

The above highlights the situation of a more congested transportation environment because of China’s current transportation strategy. Moreover, the government does not apply its transportation policy sufficiently to restrict the growth of the number of vehicles, thus allowing traffic problems to increasingly get out of control.

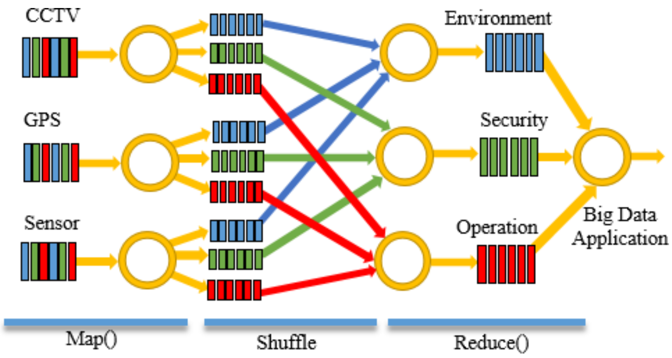

Fig. 5. Big data processing steps in transportation applications.

III. Big Data Technology for Transportation

Ben Ayed et al. have discussed different cases of the application of big data technology to transportation in their paper [11]. We summarise the presented cases in Table II. Many of these make use of commercially available technologies. Commonly used technology for data capture include: GPS for capturing location data; video surveillance and image processing technology for capturing and recognising car license plate data; traffic sensors for capturing traffic density and flow; ship sensors for capturing water quality data for navigable waterways; weather sensors for capturing current weather conditions; RFID sensors for capturing movement of RFID-tagged items in logistics; and various software technologies for processing the collected data. Data processing commonly makes use of Hadoop and related technologies: Hadoop Distributed File System (HDFS) for data storage; HBase distributed database; Hadoop MapReduce for processing data; and others.

The main steps of big data processing for transportation applications are depicted in Fig.5. These are the familiar steps of big data processing, beginning with a map operation to sort/filter input data, then a shuffle (or group) operation to redistribute the mapped data by output key, and finally a reduce operation to summarise the data on each node. Input data includes CCTV road traffic streams, GPS data of moving vehicles, traffic and weather sensors, and others. The output depends on the transportation application and may include assessments of traffic congestion, traffic flow, transportation efficiency, pollution levels, as well as forecasts of traffic and pollution trends, among others.

Technical solutions can differ in their degree of complexity. Many existing applications are very simple. For example, in Beijing (as well as in other major cities around the world) road-side electronic displays such as the one in Fig.6 show traffic conditions of nearby road sections, and have been in use for several years. Thus a driver may react immediately and choose the least congested road that leads in the general direction of travel. However, these systems only cover a small area and lack the foresight to direct drivers away from

TABLE II
CASES OF BIG DATA FOR TRANSPORTATION AND LOGISTICS (SOURCE: [11])

Country	Big Data System	Purpose	Technology
India	Real-time vehicle monitoring system	Improve operational efficiency	GPS GPRS HDFS
Ireland	Public transit system	Improve public bus transportation, reduce traffic congestion	CCTV GPS Traffic sensors
Vietnam	Traffic management system	Reduce traffic congestion and pollution	Traffic sensors Ship sensors
UK	Air travel customer analysis system	Understanding clients' needs	Smart phones Social media sites Call centre records
Sweden	Real-time intelligent transportation system	Predict future traffic conditions	GPS Traffic data Weather data
Tunisia	Logistics tracking system	Real-time cargo monitoring	IoT SaaS cloud GPS RFID HBase

Fig. 6. Beijing traffic flow sign (photo by Garnet, unmodified, licensed through Creative Commons CC BY-NC-ND 2.0, creativecommons.org).

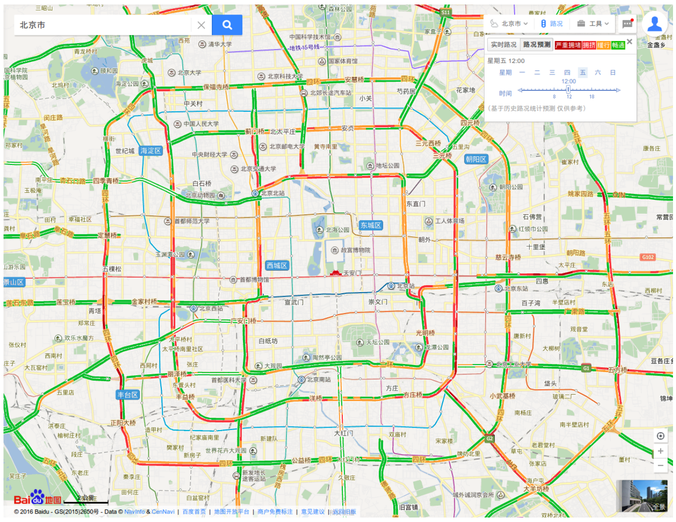

Fig. 7. BaiDu Map traffic condition forecast.

congested areas that are outside their immediate scope but in the direction of travel.

More recently geographic information systems such as Baidu Maps, Google Maps and others provide a display of the traffic situation of a larger geographic area, by aggregating traffic data collected from a multitude of roadside sensors, CCTV cameras, and even vehicle-based sensors. This allows a driver to plan an entire route based on the actual traffic congestion situation. When historic traffic data is collected and retained then this can be used to build up a profile of the traffic situation for a given time slot, such as for each hour of each day of the week. Thus a forecast of expected traffic can be provided, as illustrated in the forecast of next day's traffic in the Baidu Map of Beijing that is shown in Fig.7.

Collected historic data can also be used to reveal interesting patterns through visual analysis. A simple example of visual traffic analysis is HubCab developed at MIT, and shown in Fig.8. It allows a user to understand travel patterns, in this case patterns of taxi travel from a start to an end point. Some locations emerge as being predominantly start points, and others conversely as being predominantly end points. This has an influence on traffic flow and density, as traffic tends to flow in one direction (from start to end point), thus contributing to potentially uneven traffic density in different directions. Once such traffic patterns are discovered, they can be used by policy makers in planning longer-term improvements to the traffic network. Related work has also been done in China [12], [13].

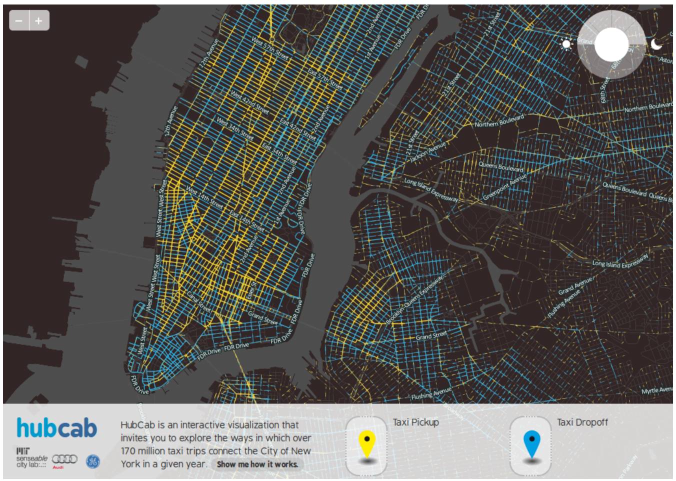

Fig. 8. HubCab big data application.

IV. Opportunities for Improvement

Having outlined the China transportation system and existing big data technology above, we identify areas in which there exist opportunities for advancing the current state of big data practice in the transportation sector in China. This is becoming an increasingly pressing issue as traffic problems in China are getting worse and the existing traffic network is reaching higher levels of saturation.

A. Big Data Standards

As the market for big data technology and its uptake in government and local traffic administration in China is still relatively immature and developing, there is very little guidance and direction given to technology providers. Coupled with the size of the country, this has resulted in great differences in intelligent transportation technology and big data technology between different provinces and regions of China. Numerous vendors provide their own technology solutions, with different functionality and implementation approach, which largely are mutually incompatible.

Efforts for making better use of data captured by existing intelligent traffic systems would benefit from a greater degree of standardisation. A few standards related to big data already exist, such as those defined by the IEEE [14]. However, there is acknowledgement that there is not enough standardisation in the field of big data [15]. China, as a large scale technology user, has the potential to develop standards for its own big data systems that may be beneficial not only to its own projects but that may define the state of the art to be used beyond its borders.

A national standards body should define a set of standards to be used by the government and local traffic administration for big data projects. These should then mandate adherence to the standards by technology vendors.

B. Standardised Software Systems

Once component-based, modular big data standards are defined, standardised software systems complying with these mandated standards can be developed. This will open the door for competitive development by various vendors of big data software system components. On the one hand the software market will become more focused, as there will be a strong incentive to work towards established standards. On the other hand the market can also enjoy diversity in different directions of big data software, such as data capture, storage, processing, visualisation and control, to name a few. When software is standards-compliant and modular it will be interoperable and interchangeable, benefiting users (i.e. governments and transport authorities) by enabling them to select best-of-class software, and to continuously upgrade and extend their systems with the components that best match their current needs.

C. National Transportation Big Data Analytics Framework

When big data standards are adopted and standardised software systems are employed, a national transportation big data analytics infrastructure can be built up. Fig.9 presents a multi-level framework for big data analytics at the municipal, provincial and national level.

At the lowest level, shown in the oval at the bottom right of Fig.9, sensors from the municipal traffic network, such as traffic cameras, radar speed detectors, GPS devices carried on vehicles, and in-road sensors, collect traffic data, and process this data in real time in the municipal analytic system. The output of this system is intelligent traffic control which switches traffic lights and traffic guidance systems in order to maximise traffic throughput and minimise congestion.

On the next higher level, shown in the oval near the centre of Fig.9, multiple municipal analytic systems feed summarised traffic data to the provincial analytic system. The provincial level analytic system is responsible for directing traffic along roads and highways between municipalities, again minimising congestion and maximising traffic throughput. This may involve feedback to municipalities, for example to have a municipality direct traffic to a less congested highway and thereby ensure a better traffic distribution within the province's traffic network. Such a higher-level decision by the province analytic system may thus provide an input to the municipal analytic system to revise its own traffic management decisions.

Similarly, at the national level multiple province analytical systems feed summarised traffic data to the top level, where national-level traffic decisions are made, that may feed back down to the provincial level where traffic management decisions may be revised, which may then have the same effect one level further down in individual municipalities. Thus the entire nation's traffic management benefits from the availability and sharing of actual traffic data at all levels, with adaptive decision making at all levels, taking the overall traffic situation into account.

D. The Future of Transportation

Extrapolating several current trends in transportation and looking one or two decades ahead, we can make predictions about the future in which big data will play an ever greater

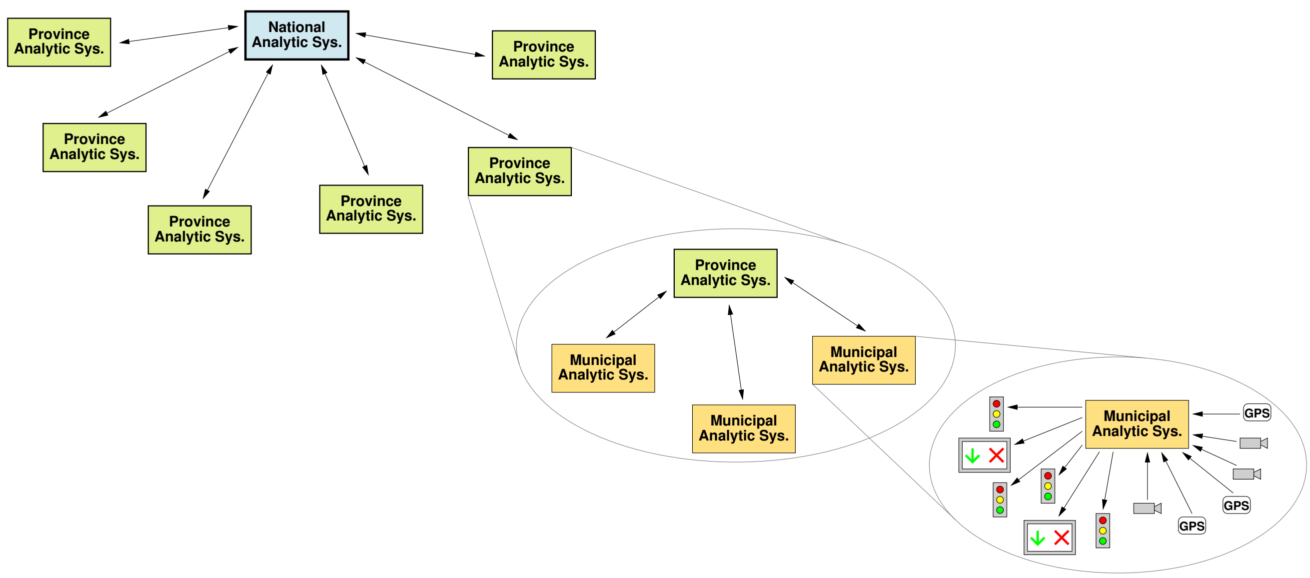

Fig. 9. National transportation big data analytics framework.

role. Driverless cars [16], [17], [18], which are already being road-tested by several research groups, are expected to increasingly become the norm. This is expected to lead to fewer accidents and improved road safety. Cars will also increasingly move away from fossil fuels and instead be electrically powered, thereby reducing pollution. Cars will become mobile communication hubs, sharing driving, weather and traffic information with other nearby cars, and with road traffic management centres. This data will be processed to enable adaptive traffic control, where traffic flow is regulated optimally by switching traffic lights based on actual traffic conditions, rather than in fixed time intervals as is currently the case. Traffic guidance will not be limited to road-side electronic traffic signs but will be transmitted directly into each car, providing each car its own individualised traffic and routing recommendations.

V. Conclusion

Data is being captured in all domains of society's life, and transportation is one of them. The large amounts of data that are increasingly detailed, fine-grained and of ever greater coverage, allow traffic and transportation to be tracked to an extent previously not possible. Existing big data analytics for transportation is already yielding useful applications in the areas of congestion management, traffic routing, and scheduling. However, we believe that this is only the beginning of a much larger penetration of applications of big data that will ultimately make the transportation network better managed, more efficient, and will identify and predict future traffic needs.

References

[1] H. Sussman, *Victorian Technology: Invention, Innovation, and the Rise of the Machine*, ser. Victorian Life and Times. Praeger Publishers, 2009.

[2] Cambridge Systematics, "Crashes vs. congestion: What's the cost to society?" AAA, Tech. Rep., Nov. 2011.

[3] OECD, *The Cost of Air Pollution: Health Impacts of Road Transport*. Paris: OECD Publishing, 2014.

[4] Ministry of Transport of the People's Republic of China, "2014 nian tie dao tong ji gong bao [2014 railroad statistics report]," http://www.mot.gov.cn/tongjishuju/tielu/201510/t20151016_1906182.html, Apr. 2015.

[5] National Bureau of Statistics of China, *China Statistical Yearbook 2015*. China Statistics Press, 2015.

[6] Intertraffic China, "Intelligent transportation industry market size will exceed 70 billion yuan," http://www.china-its.org/cnt_127.html, 2015.

[7] H. Zhang, "Beijing still struggling to deal with traffic congestion," *South China Morning Post*, Aug. 22, 2013.

[8] R.B., "How the Chinese cross the road," *The Economist*, Sep. 3, 2014.

[9] X. Zhang, H. Yao, G. Hu, M. Cui, Y. Gu, and H. Xiang, "Basic characteristics of road traffic deaths in China," *Iranian Journal of Public Health*, vol. 42, no. 1, pp. 7–15, 2013.

[10] N. Stechyson, "Bumper to bumper, all the way to Beijing," *The Globe and Mail*, Aug. 24, 2010.

[11] A. Ben Ayed, M. Ben Halima, and A. M. Alimi, "Big data analytics for logistics and transportation," in *Proceedings of The 4th IEEE International Conference on Advanced Logistics and Transport (IEEE ICALT'2015)*, May 2015, pp. 311–316.

[12] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. van de Wetering, "Visual traffic jam analysis based on trajectory data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2159–2168, Dec. 2013.

[13] M. Lu, C. Lai, Y. Tangzhi, J. Liang, and X. Yuan, "Visual analysis of route choice behaviour based on GPS trajectories," in *Proceedings of the 2015 IEEE Conference on Visual Analytics Science and Technology (VAST)*, Oct. 2015, pp. 203–204.

[14] M. Rozenfeld, "Standards that support big data," *The Institute*, p. 15, Sep. 2014.

[15] B. McKenna, "Big data needs standards to bring economic benefit, says BSI," *Computer Weekly*, Jan. 21, 2016.

[16] G. Seetharaman, A. Lakhotia, and E. P. Blasch, "Unmanned vehicles come of age: The DARPA grand challenge," *Computer*, vol. 39, no. 12, pp. 26–29, Dec 2006.

[17] U. Rasheed, M. Ahmed, M. J. Afridi, and F. Kunwar, "Road trajectory mining and autonomous steering control for vision-based unmanned vehicles," in *10th International Conference on Intelligent Systems Design and Applications*, Nov 2010, pp. 197–202.

[18] M. Beyeler, F. Mirus, and A. Verl, "Vision-based robust road lane detection in urban environments," in *IEEE International Conference on Robotics and Automation*, 2014, pp. 4920–4925.