International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 17 (2015) pp 38038-38044

Big data and transport modelling: opportunities and challenges

P. Chandrasekar
Associate Professor, Department of Civil Engineering, Velammal Engineering College, Surapet, Chennai 600066, India.
Email: chandra_traffic@yahoo.com

Abstract
This paper discusses the potential of using big data in transport modelling. In recent years, scientific research communities have been showing an increased interest in using big data; especially after technology-vendors demonstrated how effectively big data can be used to cause a significant improvement in business operations and customer experience. Personalized customer service and volume-to-value, are some of the popular phrases in big data businesses now. While this might be valid in day-to-day consumer products and services markets, the use of big data in transport research is yet to be embraced widely or yet to be documented in detail. In the past, it was the research community who were seeking suitable data for validating their models. Significant amount of resources were allocated just for the purpose of data collection alone. One example was, creating a city-wide vehicle-based origin-destination matrix. Today, big data can deliver such matrix at ease, along with numerous other travel behaviour related information. Therefore, instead of researchers seeking data, now it is common to see big data owners seeking researchers to come up with ways of utilizing the data. The question for research community is therefore this: should we re-invent the wheel of transport models that were once created with limited data available then? Or, should we re-create the models from scratch, in order to make use of an omnipotent system of data. Methodology used for this study encompasses a detailed review of recent past and current studies and articles in this field. The contribution of this paper could be an alert to stakeholders on where to focus and where not to, when it comes to injecting big data concepts in arriving at transport solutions.

Keywords: Big Data, Transport Modelling, Call Data, Smart Phone Data, Social Media Data, Analytics.

Introduction

Big data is a term that has a variety of definitions and interpretations. As big data phenomenon apparently originated from technology-vendors like IBM, Google and similar industry players (rather than academia per se), early definitions of big data are invariably available in vendor websites [1, 2]. One of the authentic definitions for big data is available from the Department of Science and Technology (DST), India, as follows: big data is data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it [3].

Big datasets were in use long before big data got popular. For example, data from census, household surveys and travel surveys, and loop detector data. However, they are collected with pre-defined purpose using systematic methods; whereas big data is generated spontaneously (and often unintentionally) by individuals on cellular and internet networks. There are no rules on types, frequency or even structure of data generated. This leads to associate big data with certain unique characteristics, namely its size (volume), its temporal component (velocity) and its inherent types (variety). Such characteristics are referred to as 3 'V's. Khan *et al* defines big data with seven 'V's (volume, velocity, variety, veracity, validity, volatility and value) [4]. The DST consolidates them into 4 'V's, namely volume, variety (structured and unstructured data), velocity (high rate of changing) and veracity (uncertainty and incompleteness). These characteristics arise due to the very nature of unconventional sources of big data such as mobile phones, online social networks, smart cards, credit card transactions or any medium where users leave their digital footprints to be traced.

Technology-vendors have demonstrated on how effectively big data can be used to improve business intelligence and consumer experience [5, 6]. The use of big data in transport research is yet to be embraced widely or perhaps yet to be published. This paper reviews notable big data applications in transport field to look for opportunities and challenges in using big data for transport modelling. For brevity, the scope of this paper is limited to land transport modelling.

Transport Modelling and Data Needs

Transport modelling is the process of formulating, calibrating and validating mathematical models that describe travel behaviour of people. Transport models deliver outputs on where, when and why people travel, and more importantly, which modes and routes they choose. Planners use such models to forecast travel demand for another twenty or thirty years, and then make justified decisions about improving the existing transport infrastructure or adding new facilities. Transport infrastructure in cities around the world now, were the result of meticulous planning carried out several decades ago using transport models available then.

Traditional urban transport planning process involves four-step modelling approach, which uses trip generation, trip distribution, mode choice, and traffic assignment models [7, 8]. There are also land-use models [9] and activity-based models [10]. Most of these models were built during 1950s and 60s. They all assume that travel is a derived demand. However, they vary from each other on the reason from where travel is derived (socio-economics, land-use, activities etc).

Trip generation models require socio-economic and travel pattern data from every person in a family or household sampled. Trip distribution models require travel time and other costs associated for making a trip between zones. They also require origin-destination (OD) matrices for checking the models built. Mode choice models require as much data as possible from various modes used (cost, waiting time etc). Traffic assignment models vary in nature and vary in data requirement as well. Dynamic traffic assignment and micro simulation models require highly disaggregated data on junction, roads, routes and networks level performances. Activity-based models require itinerary of individuals to derive travel patterns.

Data collection process is more or less the same, irrespective of the modelling approach adopted. Supply related data such as road and transit network infrastructure can be obtained from authorities and mapping agencies. It is the collection of data about people and their travel pattern, which is known to be a tedious and time-consuming exercise. Reason relies on sampling size. Even a small sample of 1 to 2 % of households in a large urban population will result in several thousand people to be surveyed through household interviews. A Singapore study involved 1% of households, resulting in 10000 home surveys [11]. Toronto study used 30000 households, little over 1% of number of households [12]. In India, Bengaluru transport modelling study engaged 2% of households (26000 interviews) [13]. Chennai study involved 38000 home interviews [14]. In addition to household interviews, a study involves a number of other surveys as well like road side interviews, travel time studies, and so on. It takes even two years to complete such data collection process [7]. With this background on what it takes to build a transport model, big data applications in transport field are further reviewed.

Big Data Applications

Transport-related big data applications can be categorized based on the source of generation of big data used. Four such popular sources are covered here: cell phone call-records, smartphone apps (using sensors), transit smart cards, and online social media. Studies from cities around the world are presented. Where available, studies from developing countries are included to relate to Indian region. The underlying opportunities and challenges are discussed later in subsequent sections.

Cell Phone Call-Records

A call-record refers to data about a phone call, and not the conversation made by people during the call. When a call is made, service-providers log data on who calls who, when do they call, and for how long they communicate etc, for billing purposes. If such data are archived for a few years, it evolves into big data. A call made by a person at two different places on a same day implies a trip had been made between those two places. Location and timestamps from archived data can be used to derive travel pattern of users. These studies do not require capabilities of a smartphone or even access to internet, making it suitable for developing countries.

As a part of research challenge called ‘Data for Development’, the European telecom operator Orange released 2.5 billion call-records from Ivory Coast, Africa in 2013 [15]. This covered data about calls and messages from 500,000 people in the city of Abidjan, during a period of five months. Anonymity of users was ensured. IBM (a participant), showed that such archived data can be used to optimize public transit operation [16]. Around 15,000 OD flows were estimated. Another group of researchers from India used Orange data to compare urban and rural travel distances and frequencies [17]. Orange had released data for Senegal in 2014 for encouraging more such studies.

A study in Israel estimated long-distance trips based on cell phone records [18]. OD tables for such trips were derived using cell phone location data. Around 10,000 users were randomly drawn every week for tracing. In 16 weeks, 80 million call locations were traced. It is interesting to note that a ‘home’ is identified for mobile phone user by the longest time the phone is linked to one particular cell (on a week day). The study reported that the number of data samples were seven-times more than that of a previous household survey conducted in late 90s.

Tracking a small number of people but for long durations helped German researchers to study stopping behaviour of travellers [19]. As a part of Nokia’s Mobile Data Challenge [20], the research work combined the Global System for Mobile communications (GSM) and Global Positioning System (GPS) data made available for 38 persons from Switzerland for over one year time. The study found that long-term GSM data is well suited to detect frequent stop locations. Other such applications using call-records data are documented by Furletti *et al* from Pisa, Italy [21], and Papacharalampous [22] and Romph [23] from Netherlands.

In Seoul, Korea, late-night service bus operations were planned based on cell phone call data [24]. Volume of calls made for taxi bookings was isolated first. Then, locations were identified from where high volumes of calls were made. Connecting such locations formed transit routes. Innovative procedures like this are proven possible with the help of telecom operators and big data analytics.

Above studies showed that OD matrices can be estimated from archived cell phone call-records. Data were gathered in a passive way, meaning without active participation or interaction from a user. However, such user interaction is vital when smartphone apps are used for travel research.

Smartphone Apps

Apps are software applications targeted for use in mobile platforms. Smartphones have one or more built-in miniature sensors such as motion sensors (for measuring acceleration and rotational forces), environmental sensors (for temperature, pressure, illumination, and humidity), and position sensors (for orientation) [25]. Developers are keen to develop apps engaging the power of sensors in the phone device. In 2014, around 1.6 million apps were available leading to billions of downloads [26]. Free and user-friendly development kits enabled anyone to write a simple app after a few days of learning. Most of the apps are entertainment, gaming and business oriented. There are many consumer-end travel assisting apps too [27]. Developing an app for transport data collection purpose is not difficult, but finding smartphone users to participate might be.

Cottrill *et al* share their experience in developing an app to conduct travel survey in Singapore [28]. The data collected will be used in activity-based models. Named as Future Mobility Survey, the app requires users to input their activity pattern. Data on mobility is generated from sensors in the phone (accelerometer etc). Users need to allow a 14-day period of monitoring their mobility sensed through the app. The study attracted only 30 participants in spite of a 25 US$ incentive. It is reported that not all participants were co-operative in providing the data required regularly. Battery drain (because of continuous use of sensors by the app) was reported as a major problem.

A study in Toronto, Canada by Abdulazim *et al* [29] used an app that combines location data from smartphones and physical movement data from accelerometer and gyroscope sensors in the device. Uniqueness of this study is that, mode of travel was automatically detected using sophisticated algorithms. The need for data entry on travel mode is avoided, which makes it easier for participants, especially during intermodal trips.

Greaves *et al* developed web-based diary (accessible via smartphone) for conducting travel and activity survey in Sydney [30]. Around 16000 trips were recorded from 847 participants. Battery life was a concern. Only 76% of participants used the portal for all the intended seven days. Young male users found the survey, in authors' term 'burdensome'. This shows that irrespective of medium used, paper or smartphone app, a successful survey must keep human mind engaged to provide credible inputs.

A pilot study by Hopkin *et al* in London involved 14 participants, to test their app [31]. Data for understanding value of time with respect to trip purpose was collected. The user interface of the app seemed to receive suggestions for improvement, before engaging a larger population to use it. Feedback showed that participants are cautious to know the purpose and final ownership of their input data. Some even suggested incentives like prize draw, a remainder to show community participation may not be guaranteed just because the medium of data collection is new and novel.

Berkeley researchers Gopal *et al* conducted a pilot study in Pune, India, using their app to collect per-second data on car driving behaviour [32]. Data was then used in Berkeley lab for fuel and engine efficiency analysis. Three persons were requested to install the app to gather data for a few days. The study is unique in considering real-world vehicle performance characteristics that could form part of inputs for mode choice models.

Some of other studies using smartphone apps are: Deutsch *et al* [33], Fan *et al* [34] and Misra *et al* [35], covering travel behaviour and crowd-sourcing aspects. Crowd-sourcing is community participation to generate data and share information for fellow travellers. One such crowd-sourcing app in India is note-worthy (though not from academic arena) [36]. It is a big data-based web and mobile phone app providing insights into the Indian railway network and train operations. Archived data about 50 million train travellers across 8,000 stations and 2,000 routes in India were used and analysed. Location data from mobile phone users, who are travelling in trains, are used to provide real-time results for anyone using the app. This demonstrates the possibility of community participation, if the application is useful for population at large.

Transit Smart Cards

Fare collection using smart cards results in time savings and convenience. The role of smart cards as a big data generator was realized only recently. Every time a passenger boards to or alights from a bus, and enters or leaves a train station, there is digital footprint of the person, with data on origin and destination of the transit trip and timestamps. The following shows successful applications of smart card data.

As a part of evaluating master plan for Beijing, China, Yu *et al* investigated the effectiveness of subway train infrastructure [37]. A single day's data was used, which turned out to be 8 million passenger trips. Combined with GIS system, researchers compared the observed and expected travel behaviour. The study enabled planners to identify areas that developed as per expectation and also areas which weren't.

In the city of Transantiago (capital of Chile, South America), smart card data was combined with bus location data from GPS [38]. Around 6500 buses fitted with GPS equipment generated 80 to 100 million records per week (recording location every 30 seconds). Smart cards generated 35 to 45 million transactions per week from boarding and alighting at 10000 bus stops. Big data analytics provided insights on load profile along routes, OD flow data of passengers and speed profile of transit fleet.

The Bus Rapid Transit (BRT) system in Istanbul, Turkey, stretches for around 50 km with 44 stations. Gokasar *et al* used smart card data from the system amounting to 6 million transactions per day, to study transit operational performance [39]. Waiting times were also analyzed considering the time gap between users tapping their cards at station entry and the time of arrival of buses. Time-dependent OD tables and travel times were also derived and analyzed.

Singapore's Land Transport Authority announced plans to analyze public transit operations in collaboration with IBM, in 2014 [40]. While smart card data will provide the number of commuters entering and leaving the transit stations, telecom operators will provide data on origins and destinations. This indicates that access to stations (by walk or other modes) is also considered besides analyzing transit operations.

One important observation in relation to smart card data generation is that, it resembles cell phone data generation, with both methods being passive in nature. Users are not required to participate or interact to provide any inputs. Digital footprints of users are traced and used instead. This eliminates limitations involved in user interactions, common in smartphone app cases discussed before, and social media cases to be discussed next.

Online Social Media

There are around 200 online social media networks and micro-blogging web-sites and services available [41]. At least a dozen of them have 100 million or more active users, while Facebook and Google have more than billion members. Twitter and WhatsApp have more than 500 million members. There were studies relating social media network and travel behaviour even before big data became popular. Sharmeen *et al* provides a review of such works [42]. The following however, are recent works.

Unstructured data originate in large quantities at almost real-time speed, when millions of people share messages, photos, audio and video files and chat about them, any time of the day from any part of the world. Mining such data provides a broad sense of what people feel in general. A popular application in this type is sentiment analysis [43]. Data is analysed and classified under three categories of expression or feeling: positive, negative or neither. Such sentiment analysis on social media data reveals people's opinion about any issue, including transport performance, real-time.

A recent study in Singapore compared Twitter messages (tweets), smart card data and household interview survey [44]. Tweets were geocoded and hence location data were available. Clustering algorithms were applied as a part of analytics. The study was conducted for entire Singapore city, for around six months. Around 2 million tweets were used. Good correlation was found between social media data and other forms of data (with correlation coefficients 0.7 and above). Social media data were available at a very low cost when compared to household surveys.

Xerox Research Centre in New York engaged Binghamton University researchers with 15 million tweet messages along with time and location data [45]. Travel pattern of around 270,000 people were studied. Findings about how people often change their preferences in choosing restaurants were unique. This shows that destination choice for leisure trips gets reflected in social media data. It is common to see users sharing their current whereabouts, especially if the trip is for eating, shopping or entertainment purposes. Such data are useful in understanding short intra-zonal trips and mode choices associated with it.

Sentiment analysis helped researchers to know the mood of even an entire city. Bertrand *et al* had developed algorithms to analyze tweets to know what kinds of feelings are expressed (including messages with emoticons like smiley symbols) [46]. Entire New York City was sampled for tweets and a high resolution 'mood map' was created. Interestingly, traffic and transport related spots showed a strong negative sentiment, possibly a reflection of traffic delays. Happiest feelings were from public parks. Though it is not possible to know reasons of mood directly, an extreme negative feel could alert concerned authorities.

There are also virtual locations sharing services like Foursquare, where members can virtually 'check-in' at different venues (shops, hotels, etc) and post their comments [47]. Members can also interact with other members. Cheng *et al* investigated 22million check-ins across 220,000 users and analyzed them for human mobility patterns [48]. Analysis indicated that users are more likely to express complaints or negative sentiment. Users also share mobility information like where from they travelled and other related information. Destination choice models can benefit from such inputs.

There are also initiatives that cover various technologies within one large project. EUNOIA, a research project in Europe examines big data and mobility in the context of smart cities where online social network data are combined with other sources such as smart cards, mobile phones and credit cards [49]. Barcelona, London and Zurich are participating cities. United Kingdom is also into Intelligent Mobility projects [50].

As big data is relatively a new paradigm compared to transport modelling that is several decades old, only well documented cases were presented above. Next section discusses the underlying possibilities and hurdles for transport researchers.

Opportunities and Challenges

Both opportunities and challenges are discussed together for clarity. From a previous section on 'transport modelling and data needs', it can be realized that transport models are data-hungry. Transport modelling attempts to mathematically mimic one of the most complex behaviours known, which is, humans responding to choices. Scientists and researchers from the mid-20th century achieved triumph upon developing travel forecasting models with bare minimum computing facilities available then. Previous section on 'Big Data Applications' showed that the present-day researchers are provided with access to any data potentially, at any required level of detail. Big data appears to promise such omnipotence on the first impression. However, if one gets to know about various 'V's (volume, velocity etc) attached to it, challenges arise; but not necessarily to transport modellers, but unquestionably to computing and data scientists. Most of the publications reviewed in previous section were authored by four or more persons, with at least one of them from computing or data science field. Multi-disciplinary and cross-disciplinary works can therefore be seen increasing in transport research.

Transport modellers know what data are required in order to describe travel behaviour mathematically. If big data can provide such data, it will save time and resources spent on household interviews. Big data by default is not something generated to serve travel data collection (unless apps are written to generate data for such purpose). For example, consider archived cell phone calls data. People make calls for various reasons; what is new is the possibility for telecom operators to save data about every call made by every person registered to them. There were around 900 million mobile phone users in India in 2014 [51]. Urban areas have around 500 million users. Even two calls per day can generate 1 billion records every day. There was no way to store and process such amounts of data a decade ago. Big data applications reviewed earlier showed that it is a possibility now.

Cell phone data based studies showed that OD matrices can be generated from archived data. OD matrices are usually the output of third step in four-step process, at the expense of significant amount of resources. A cell phone call-record generated for billing purposes becomes useful to transport modeller, due to big data technologies. With suitable sensor apps, even the mode of travel was also shown detectable automatically. It is also possible to know the type of trip end (home or office etc) by mining text messaging data from social media. The challenge for a data scientist is therefore to develop algorithms for filtering or extracting what transport modellers require. Better the analytics tools and techniques, easier and faster will be the process of mining the data to extract inputs for existing transport models.

Smartphone data studies proved that activity patterns can be logged with user interaction with apps. Such data are vital in activity-based modelling of transport. However, it is disappointing to see the number of users who can volunteer to be traced, possibly due to privacy concerns (even after providing cash incentives in some cases). Therefore, smartphones apps can be useful in data collection, only if people volunteer. Instead of cash, other ways of incentives can be thought of, such as free talk time, which is proven successful in business applications [52].

Other challenges associated with smartphone apps are sampling and security. A study cannot fully rely on smartphone data alone, as smartphones users represent one distinct section of population only: those who can afford it. Not everyone in a household can be expected to own a smartphone and not all smartphones can be assumed to have sensors required for transport study. On security aspect, apps are vulnerable to malware, virus and hacking [53]. Data gathered should be made sure that they are from intended users (and not hackers). Travel surveys or transport studies using apps should take into account such limitations and vulnerabilities.

Compared to smartphone apps, big data from public transit cards seems to be much more credible and authentic reflection of transit demand (as there is no need for direct user interaction). Transit authorities usually provide data on routes, stops and stations, and transit usage on aggregated levels. Smart card data now provides snapshots of public transit usage throughout the day at a highly disaggregated level. Smart card data provide opportunity for enhancing mode choice models and transit assignment models.

While smart cards provide passive data, social media provide a rather highly active data. Social media data seems not readily useful for providing traditional transport modelling with required data, but it can be used at evaluation stages for faster and direct opinion from the public. However, one should be aware of increasing fake accounts and bot-accounts [54]. A 'bot' is like a software robot [55]. They are programs that imitate human accounts in a social media network. Bots are known for malicious activities. Bots can mislead polls. Besides bots, another credibility issue arise from the involvement of children in social media. Often, children using game consoles (like Xbox etc) interact with fellow gamers online via twitter [56]. Filtering sentiments that are not expressed from unintended users such as children and bots is necessary.

Nevertheless, considering an ever-increasing smartphone usage and social media penetration in urban life, new data analytics driven models are needed to capture travel behaviour. Treating traffic as fluid had its time in history. It was novel at its time. With unlimited ingenuity of human mind, it is possible for transport modellers to use the big data opportunities to re-invent the wheel of transport modelling. Mood maps and sentiment research showed how researchers can read people’s mind. New models that are not only disaggregate but also fully flexible to major changes in technology are needed to be built. Because, technologies such as vehicle-to-vehicle communication (V2V) [57] and internet of things (IoT) [58] will generate even bigger amounts of data in real time, and it is unknown how travel behaviour will be influenced.

Summarising it all, it is evident that archived data of cell phone calls and smart card data are useful in extracting OD matrices and route choices. They can be of much use in popular transport modelling approaches. Smartphone apps and social media are yet to be proven successful in large-scale research. In India, two ambitious plans are on the way, one with building 100 smart cities and the other with plans to rejuvenate 500 existing cities [59]. Indian telecom operators could share phone call-records to develop algorithms that suits to local applications (following what Orange does in Europe).

Conclusion

Big data applications were reviewed to find whether big data can be of use in transport modelling. While most of the available literature discuss about what can be done potentially, only a handful of studies do report what had been done successfully. Such works were reviewed under four mediums of data generation: cell phone call-records, transit smart cards, smartphone apps and social media. It is found that archived cell phone call data from telecom operators and smart card data from transit operators are more useful than smartphone apps and social media data. Note that cell phone call data does not refer to what people speak or communicate; it refers to where and when calls were made, thereby location during the call can be traced. If call locations vary, it implies a movement or trip. Smartphone apps in research arena are found not as successful as they do in consumer markets, because of privacy concerns of being monitored. Social media data seems suitable for short-term quick evaluation purposes only, by methods like sentiment analysis. As such, in the context of developing economies like India, archived cell phone usage data can be very useful to transport modellers. If data on trip generation and origin-destination matrices are extracted from call-records using big data analytics, the findings can then replace household and travel surveys for smaller short-term studies and augment data for larger long-term studies. As India is about to begin projects on 100 smart cities and 500 redevelopment cities in the near future, opportunities arise for using traditional models as well as building new models that can derive new insights from big data. Indian telecom operators could share such big data on call-records (like Orange in Europe) to engage researchers in developing new algorithms that suit to local applications. Other smart technologies can be applied later for evaluation and monitoring purposes (as smart infrastructure will be available then). Data owners (telecom operators, technology-vendors, social media companies, transit operators and government agencies) could find a win-win situation if big data is shared and used in enhancing transport modelling and planning process, thereby creating an economy which is beneficial to all stakeholders, including data owners themselves.

References

[1] http://www-01.ibm.com/software/in/data/bigdata/
[2] http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf

[3] http://dst.gov.in/scientific-programme/bigdatainitiative.html
[4] Khan, M. A., Uddin, M. F. and Gupta, N., 2014, "Seven V's of Big Data - Understanding Big Data to extract Value". Zone 1 Conference of the American Society for Engineering Education, United States.
[5] http://www.sap.com/solution/big-data/software/applications.html
[6] http://www.retailcustomerexperience.com/articles/how-retailers-can-turn-big-data-into-real-insights/
[7] Kadiyali, L. R., 2014, Traffic Engineering and Transport Planning, Khanna Publishers, New Delhi.
[8] Ortuzar, J. D. and Willumsen, L. G., 2011, Modelling Transport, John Wiley and Sons Ltd, United Kingdom.
[9] http://www.casa.ucl.ac.uk/rits/lowry.pdf
[10] Malayath, M. and Verma. A., 2013, "Activity based travel demand models as a tool for evaluating sustainable transportation policies", Research in Transportation Economics, Vol. 38, Issue 1, pp. 45-66.
[11] http://www.inrosoftware.com/assets/Model-City-2014-Presentations/04-03-Wee-Liang-Lim.pdf
[12] http://www.nymtc.org/project/surveys/workshop/Rousseau1.pdf
[13] http://embarqindiahub.org/sites/default/files/Comprehensive%20Traffic%20and%20Transportation%20Plan%20for%20Bangalore.pdf
[14] http://www.cmdachennai.gov.in/pdfs/CCTS_Executive_Summary.pdf
[15] http://www.orange.com/en/home
[16] http://www.komindesign.com/oma/big_data_2/IBM_Research.pdf
[17] Yadav. K., Kumar. A., Bharti, A. and Naik, V., 2013, "Technical Paper on Characterizing Mobility Patterns of People in Developing Countries using Their Mobile Phone Data", Indraprastha Institute of Information Technology, Delhi, India, https://repository.iiitd.edu.in/jspui/bitstream/123456789/109/1/IIITD-TR-2013-002.pdf.
[18] Bekhor, S., Cohen, Y. and Solomon, S., 2013, "Evaluating long-distance travel patterns in Israel by tracking cellular phone positions", Journal of Advanced Transportation, Vol. 47, John Wiley and Sons Ltd, United States.
[19] http://research.nokia.com/sites/default/files/public/mdc-final458-schulz.pdf
[20] http://research.nokia.com/page/12000
[21] Furletti, B., Gabrielli, L., Renso, C. and Rinzivillo, S., 2013, "Analysis of GSM calls data for understanding user mobility behaviour", IEEE International Conference on Big Data, United States.
[22] Papacharalampous, A. E., 2014, "Aggregated GSM data in Origin Destination Studies", Masters' thesis, Technical University of Delft, Netherlands.
[23] Romph, E., 2013, "Using big data in transport modelling", Data & Modelling Magazine, Issue 10, Summer Issue, United Kingdom.
[24] http://policytransfer.metropolis.org/case-studies/owl-bus-based-on-big-data-technology
[25] http://developer.android.com/guide/topics/sensors/sensors_overview.html
[26] http://www.statista.com/topics/1002/mobile-app-usage/
[27] http://www.timeout.com/travel/features/1169/the-worlds-50-best-travel-apps
[28] Cottrill, C., Pereira, F., Zhao, F., Dias, I., Lim, H. B., Ben-Akiva, M. and Zegras, C., 2013, "Future Mobility Survey: Experience in Developing a Smart-Phone-Based Travel Survey in Singapore", Transportation Research Record 2354, pp. 59-67.
[29] Abdulazim, T., Abdelgawad, H., Habib, K. and Abdulhai, B., 2013, "Using smartphones and sensor technologies to automate the collection of travel data", Transportation Research Record 2383, pp. 44-52.
[30] Greaves, S., Ellison, A., Ellison, R., Rance, D, Standen, C., Rissel, C. and Crane, M., 2014, "A web-based diary and companion smartphone app for travel/activity surveys", 10th International Conference on Transport Survey Methods, Leura, Australia
[31] Hopkin, J., Ball, S. D., Hannay, A., Hutchins, R., Palmer, D., Rahman, S., Naberezhnykh, D. and Longhi, D., 2014, "Pilot trial of a transport crowdsourcing smartphone app: final report", PPR-719, Transport Research Laboratory, United Kingdom, http://www.trl.co.uk/umbraco/custom/report_files/PPR719.pdf
[32] Gopal, A., Schewel, L., Saxena, S. and Phadke, A., 2013, "The Transportation Leapfrog: Using smartphones to collect driving data and model fuel economy in India", Ernest Orlando Lawrence Berkeley National Laboratory, United States, https://ies.lbl.gov/sites/all/files/lbnl-6293e_pdf_1.pdf
[33] Deutsch, K., McKenzie, G., Janowicz, K., Li, W., Hu, Y. and Goulias, K., 2012, "Examining the use of smartphones for travel behaviour data collection", 13th International Conference on Travel Behaviour Research, Toronto, Canada.
[34] Fan, Y., Chen, Q., Liao, C. F. and Douma, F., 2013, "UbiActive: a smartphone-based tool for trip detection and travel-related physical activity assessment", Transportation Research Board 92nd Annual Meeting, United States.
[35] Misra, A., Gooze, A., Watkins, K., Asad, M. and Dantec. C., 2014, "Crowd sourcing and its application to transportation data collection and management", Transportation Research Record 2414, pp. 1-8.
[36] http://www.railyatri.in/insights
[37] Yu, W., Mao, M., Wang, B., Liu, X., 2014, "Implementation evaluation of Beijing urban master plan based on subway transit smart card data", 22nd International Conference on Geoinformatics, Kaohsiung, Taiwan.
[38] http://www3.imperial.ac.uk/newsandeventspggrp/imperialcollege/engineering/cts/eventssummary/event_2-5-2014-23-20-35

[39] Gokasar, I., Simsek, K. and Ozbay, K., 2014, "Using Big Data of Automated Fare Collection System for Analysis and Improvement of BRT-Bus Rapid Transit Line in Istanbul", Transportation Research Board 94th Annual Meeting, United States.
[40] http://www.futuregov.asia/articles/singapore-government-and-ibm-to-improve-public-transport-with-big-data-analytics
[41] http://en.wikipedia.org/wiki/List_of_social_networking_websites
[42] Sharmeen, F., Arenitze, T. and Timmermans, H., 2010, "Modelling the dynamics between social networks and activity-travel behaviour: literature review and research agenda", World Congress on Transport Research, Lisbon, Portugal.
[43] http://en.wikipedia.org/wiki/Sentiment_analysis
[44] Eggermond, M., Chen, H., Erath, A. and Cebrian, M., 2015, "Investigating the potential of social network data for transport demand models", Transportation Research Board 95th Annual Meeting, United States.
[45] Digital Bonanza – Cover story, 2014, Binghamton Research Magazine, Winter Issue, pp. 12-19.
[46] Bertrand, K. Z., Bialik, M., Virdee, K., Gros, A. and Bar-Yam, Y., 2013, "Sentiment in New York City: A High Resolution Spatial and Temporal View", New England Complex Systems Institute, Cambridge, United States, http://www.necsi.edu/research/social/newyork/
[47] https://foursquare.com/
[48] Cheng, Z., Caverlee, J., Lee, K. and Sui, D., 2011, "Exploring Millions of Footprints in Location Sharing Services", Fifth International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media, Barcelona, Spain.
[49] http://eunoia-project.eu/doc/about/
[50] https://ts.catapult.org.uk/intelligent-mobility
[51] http://www.trai.gov.in/WriteReadData/WhatsNew/Documents/Press%20Release-TSD-Mar,14.pdf
[52] http://www.rechargeoverload.in/2014/12/top-10-highest-paying-android-apps-for-free-mobile-recharges-talktime.html
[53] Hypponen, M., 2006, "Malware goes mobile", Scientific American, November issue, United States, pp. 70-77.
[54] http://www.dailymail.co.uk/sciencetech/article-2722677/Rise-Twitter-bots-Social-network-admits-23-million-users-tweet-automatically-without-human-input.html
[55] Ferrara, E., Varol, O., Davis, C., Menczer, F. and Flammini, A., 2015, "The Rise of Social Bots", Cornell University Library, United States. http://arxiv.org/abs/1407.5225
[56] http://www.pcworld.com/article/2843966/xbox-one-november-system-update-brings-custom-backgrounds-twitter-integration.html
[57] http://en.wikipedia.org/wiki/Vehicular_communication_systems
[58] https://www.ida.gov.sg/~/media/Files/Infocomm%20Landscape/Technology/TechnologyRoadmap/InternetOfThings.pdf
[59] http://indianexpress.com/article/india/india-others/100-smart-cities-project-gets-cabinet-nod/

Note: All references to websites were last accessed and checked for their availability on the 1st of September, 2015.