Towards Intra-Vehicular Sensor Data Fusion

Paulo H. L. Rettore, Bruno P. Santos, André B. Campolina, Leandro A. Villas[1], and Antonio A. F. Loureiro
Computer Science Department
Federal University of Minas Gerais, Brazil
Email: {rettore, bruno.ps, andre.campolina, loureiro}@dcc.ufmg.br
[1]Computer Institute
[1]Campinas State University, Brazil
[1]Email: leandro@ic.unicamp.br

***Abstract*— Urban mobility aspects have become a challenge with the constant growth of global population. In the same time, more data has become available, which allows new information technologies to improve the mobility systems, especially the transportation system. Thus, a low cost strategy to handle these issues, rises as a new concept named ITS – Intelligent Transportation Systems. These systems depend on various data types and sources, and aggregating it is an important task, which can be accomplished by performing heterogeneous data fusion. In this work, we conducted an exploratory analysis over real vehicular data to show for each listed data issues (i.e imperfection, correlation, inconsistencies, among others) which of them have been found in our data set. Indeed, we found out several issues in the data implying that they must be treated before fusion process. As future extensions of this work, we will apply heterogeneous data fusion techniques to enhance, for example vehicular mobility traces by adding contextual information such as traffic conditions and driver behavior.**

I. Introduction

The world's population has increased and of city dwellers has surpassed 50% of the whole population. In this scenario, huge cities have emerged and also several human mobility issues such as traffic and transit. At the same time, massive volumes of data have become available, which enabled new information technologies that can be used to improve the mobility systems.

This paper examines the vehicular sensor data aspects in Intelligent Transportation System (ITS) context. We show challenges, useful data, as well as some methods to handle issues related to the data. In particular, our focus is in heterogeneous data fusion using intra-vehicle sensor data by collecting it from the Engine Control Units (ECU) of a car. Although several papers present reviews of heterogeneous data fusion [1], [2], [3] or data fusion in ITS [4], our work provides the reader an illustration of the listed data fusion aspects with examples based on the conducted case study.

The rest of this work is organized as follows: Section II presents background information about ITS and heterogeneous data fusion; Section III describes the vehicular data used in this study; in Section IV we present our case study and illustrate the issues regarding the fusion of the data collected; finally, in Section V we conduct a discussion about heterogeneous data fusion using vehicular sensor data and present our conclusions.

II. Background

This section presents fundamental points about ITS, heterogeneous data fusion, challenges and opportunities in the field. We will start by contextualizing ITS.

Over the years, huge cities have had significant issues related to transportation and traffic, for the simple reason that people constantly need more quick and safe mobility. The real scenario shown in [5] alerts that urbanization increases congestion around the world, causing hours wasted in traffic. The Global status report on road safety 2015, says that total of traffic deaths is 1.25 mi per year. In Brazil, only São Paulo had more than 7000 deaths in 2014 and its car fleet grew up to 26 million [6]. Injuries and accidents also rose significantly, that implies in increasing of costs with medical expenses, productivity and property damages [7]. In the last years, governments have adopted strategies to provide an efficient and an additional capacity of traffic in these huge cities. For instance, traffic restriction on selected regions and rotating vehicles. Nowadays, this no longer seems to be economically or socially feasible, highlighting the need for intelligent ways to handle transportation system issues.

The transportation systems are a means of moving persons and objects (animate or inanimate), but regarding ITS, there is no standard definition, thus the term can be interpreted broadly and strictly way [8]. In few words, ITS aim to improve decision-making. ITS integrates information, communication technologies and transportation system into applications and services to efficiency boost of the transportation systems and mitigate its issues.

Any ITS instance conducts one or more the following intuitive steps: collection, processing, integration and providing information. ITS includes at least four subsystems [8], [4]: i) Advanced Transportation/Traffic Management Systems (ATMS) to control and manage traffic devices (signals, monitoring and safety devices etc...), manage emergency situations, and other apparatus that support the system; ii) Advanced Traveler Information Systems (ATIS) to collect data and process it to improve understanding of traffic conditions and derive indicators to support traveler guidance; iii) Automatic Incident Detection (AID) to apply algorithms for automatic incident detection as soon as possible to increase safety and reduce users perception of traffic disruption; iv)

Advanced Driver Assistance Systems (ADAS) to apply technologies in transportation system components (e.g. vehicles and roads) to reduce accidents and improve safety of the users, for instance, ADAS cover collision avoidance and drive assistance. Also, ITS involves others systems such as Network Control, Traffic Demand Estimation, and Forecast.

The demands of precise traffic information is an increasing challenge for public administrators and private businesses conducting to the emergence of Intelligent Transportation System [9]. ITSs subsystems are powered by data as much as possible. Traditional traffic sensors, usually, are installed to measure traffic flows at given point, but alone they are ineffective. However, other data sources are spreading, such as cameras, GPS, smartphones and probe vehicles. All these multiple sources may provide complementary data and can be used to extract more comprehensive and detailed information about the traffic conditions. Thus, providing timely and precise traffic information, allowing ITS to keep awareness about traffic status and manage processes and services built to optimize the efficiency and safety of the transportation systems.

Data information is the heart of ITS. Indeed, there is no way to build ITS subsystems without data analysis. Usually, the data is heterogeneous (such as cameras, GPS, smartphones tracking, and probe vehicles). Thus, heterogeneous data fusion techniques are suitable in such situation [1]. There are frameworks and models to perform data fusion in the literature as well as discussed in [1], [2], [3]. To perform data fusion usually are three approaches: statistical, probabilistic and artificial intelligence [4].

There are several issues that make data fusion a challenging task, especially those regarding heterogeneous data. The majority of the issues arise from the data to be processed and fused. Data fusion aspects are extensively discussed in [3]. For the authors, the data are naturally imperfect by suffering conversions (analogical/digital) or associations with some degree of uncertainty. For instance, the data presents inconsistencies due to outliers or ambiguous data obtained from sensors. Also, data can be misaligned, requiring calibration before being fused. By being part of our scope, we return more broadly in Section IV.

III. Vehicular Data

Modern vehicles rely heavily on data acquired through embedded sensors to improve the quality of their control systems. In order to better control the vehicle's behavior, manufacturers invest both on quantity and quality of the sensors they use [10]. Some of the sensors embedded in a modern vehicle include throttle pedal position, fuel pressure and oil pressure. The sensors on a car communicate with the Engine Control Unit (ECU) through an internal wired network [11] and the data they output is accessible using the On-Board-Diagnostics (OBD) interface.

The most recent OBD interface is OBD-II, and it was introduced in the 1990's to standardize aspects of the system like the physical connector (displayed in Figure 1), its pinout, the signaling protocols and format of the messages. The system is usually, employed to monitor and regulate gases emissions and must be present in all cars produced since 1996 in Europe, United States, and 2010 in Brazil. The OBD interface is also used to help aftermarket maintenance, since it provides access to engine fault codes that informs mechanics about failures on the whole car, saving valuable time when tracing the origin of the problems [12].

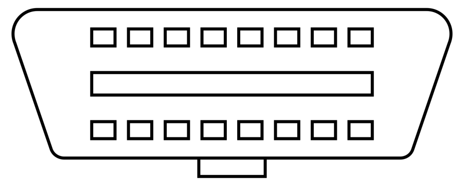

Fig. 1. Female OBD Connector

There are five signaling protocols allowed on OBD interface, as shown in Table I. All these protocols use the same OBD connector, however the pins are different except for those that provide power supply.

TABLE I
OBD Signaling Protocols

Protocol	Transfer Rates
SAE J1850 PWM	41.6 kbit/s
SAE J1850 VPW	10.4 kbits/s
ISO 9141-2	10.4 kbits/s
ISO 14230 KWP 2000	10.4 kbits/s
ISO 15765 CAN	250 or 500 kbits/s

The data collected from the sensors in the car are available through OBD Parameter IDs (PIDs). In Table II, we show some of the sensors whose readings are available using the combination of OBD and smartphone. There are also other hundreds of sensors that can be accessed using OBD's parameter ID's - some of which are defined by the OBD standard and others are defined by the manufacturers.

A. Heterogeneous Data

Even though data collected from sensors embedded in a vehicle come from the same entity - the vehicle itself - it should not be considered homogeneous. The information is collected from different sensors spread across different parts of the vehicle's body in different measuring units. The heterogeneity of vehicular sensor data does not mean that there aren't relationships between the readings of different sensors, since all of them monitor the same entity.

It is also possible to extract contextual information from data acquired by vehicular sensors. For instance, observing a car's speed over time, the traffic condition on its location can be inferred based on aspects like average speed and time stopped. These aspects represent peculiarities of traffic jams, where the average speed is low and most vehicles are stopped for long periods.

IV. Problems of Heterogeneous Data Fusion: Case Study

Khaleghi et al. [3] presented a comprehensive study of methodologies that aim at solving problems related to het-

TABLE II

SENSORS COLLECTED FROM OBD AND SMARTPHONE

Sensors					
Engine load	Vehicle speed	Torque sensor	Fuel pressure	Oxygen sensors	Fuel Tank Level
Kilometers per litre	Intake air temperature	Ambient air temperature	Catalyst temperature	Relative throttle position	Accelerator pedal position
Fuel flow rate	CO2	Ethanol fuel %	Engine oil temperature	Fuel injection timing	O2 sensor monitor
Voltage	Distance traveled	Fuel remaining	Fuel rail pressure	Hybrid battery pack remaining life	Evap. system vapor pressure
Engine RPM	Engine coolant temperature	Fuel type	Malfunction indicator lamp	Exhaust gas recirculation error	Mass Air Flow Sensor
Altitude	GPS location	Collision sensor	Automatic brake actuator	Steering angle sensor	Rear camera
GPS speed	Gyroscope	luminosity sensor for headlights	Active park assist	Water in fuel sensor	Airbag sensor
Barometric Pressure	Acceleration	Cost per mile/km	Front object laser radar	Night pedestrian warning IR sensor	Tire pressure sensor
Microphone sensor	Pressure sensor	Drowsiness sensor	Shock sensor	Rain-Sensing Windshield Wipers	Motion sensor

erogeneous data fusion. The problems are more related to the data than to the methodologies used to fuse them, mostly because data used in fusion is often collected from sensors and data sources that are inherently imperfect. The authors elaborated a taxonomy of data fusion aspects described as follows. Besides that, we present a case study to provide a practical and comprehensive data analyses in vehicular sensor data.

We considered as a case study the sensors data collected from vehicles and its relationship. We used an OBD Bluetooth adapter to collect data from a car. The logs of this vehicle consist of 55 trips of 40 km with an average time of 50 minutes each. Hereafter, we show the categories of fusion problems in a practical view. Thereunto, we choose an examples observed during the data collected from the vehicles, as our initial work.

A. Granularity

Granularity is related to the ability to derive valuable information about entities of interest on a data set. It is a concerning aspect on data fusion, specially when dealing with rough sets, when neither fine and coarse grained information is beneficial for the final process. A fine-grained information will not take advantage of the rough set techniques, on the other hand, a coarse grained data may not be enough to derive useful information.

To characterize the granularity problem in vehicular sensor data, we investigate traces of taxis, buses, cars, and their respective time interval of data collection. In the literature, it is usual to find traces with measure between each 10 and 60 seconds. Thus, we measure the speed of a vehicle from its ECU each second and GPS speed each minute. The Figure 2 shows an example of a car trace along almost 40 minutes, the Figure 2(A) and (B) present the speed vehicle and GPS speed, respectively. The Figure 2(C) shows GPS speed measured every minute. It is noted that, in the Figure 2(A) the vehicle speed is represented as fine-grained, hence more detailed vehicle behavior is perceived. For instance, looking the begin and end of the trace, it is clear to observe the stops-and-gos. This information reveals a particular behavior in a specific environment, as urban area. On the other hand, the Figure 2(C) represent the GPS speed in coarse-grained, hence it can not address the same behavior mentioned before.

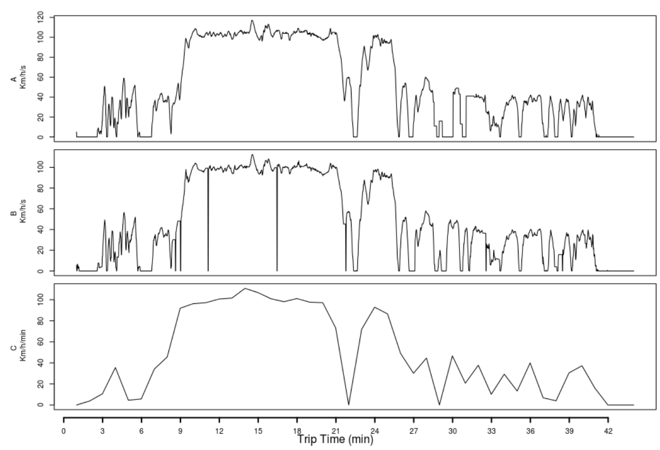

Fig. 2. Comparison Between Vehicle Speed and GPS Speed Collected Every Second and Every Minute

B. Vagueness

Vagueness occurs in data sets where attributes are not well defined. The loose definition of attributes allows subjective measures, *i.e.*: ”fast” or ”slow”. Fuzzy sets are a way to remove the subjective aspect of these parameters.

The vagueness in a vehicular data context may be intend as the speed of vehicle. In other words, it is not well defined the speed, ”fast” and ”slow”, of the vehicle. For instance, in the Figure 2(A), the highway environment is characterized by the vehicle’s speed behavior, which does rise above 80 km/h and below 120 km/h. Thereby, 80 km/h speed can be slow in a highway environment, but fast in the urban environment, where the vehicle’s speed behavior does not rise above 60 km/h, due to legislation and traffic density.

C. Outlier

Outliers are points whose data is incorrect. These situations are often caused by errors in the sensors that generate it. False data is dangerous to data fusion systems, mainly because it leads statistical inferences to imprecise results.

The environment perception from sensors may come with incorrect data. These data represent points that distorter among the major data collected. The Figure 2(B) shows the GPS speed along the trace. However, it is noted some distorter points with 0 (zero) values between high values collected. For instance, approximately in 10 minutes the values are around 100 km/h and instantly changes to 0 km/h, returning to 100 km/h after that. Similar occurrences are shown along the trace and are called outliers.

D. Conflict

A same phenomenon, when observed by two or more sensors or specialists should be perceived in the same way by all of them. However, divergent specialists' opinions or punctual errors in sensor readings happen and cause conflicts in data observations. A simple, yet questionable, conflict solution is the Dempster combination rule [13].

In the Figure 2, the conflicts appears when two sensors are related to describe the speed of vehicle. The Figure 2(A) shows, approximately, in 10 minutes the values are around 100 km/h speed. However, in that same time interval, the Figure 2(B) shows 0 km/h speed. The challenge of this topic is which one may be considered to the data fusion.

E. Incompleteness

Incomplete data is, intuitively, data with missing parts. These missing parts may lead to incorrect conclusions based on the data and, thus, must be addressed. A solution to deal with this type of data is to treat the data in a probabilistic way.

The log used in our case study was obtained using an OBD Bluetooth adapter and a smartphone. However, interferences among electronic devices inside the vehicle, or barriers in the environment as tunnels, sometimes, cause the loss of communication. Consequently, gaps are introduced in a trace and made the data set incompleteness as showed in Figure 3. The Figure 3(A) shows the vehicle speed collected from ECU and the Figure 3(B) shows in three different moments, gaps caused by interruption of communication, ignoring important information and making the results inconsistent.

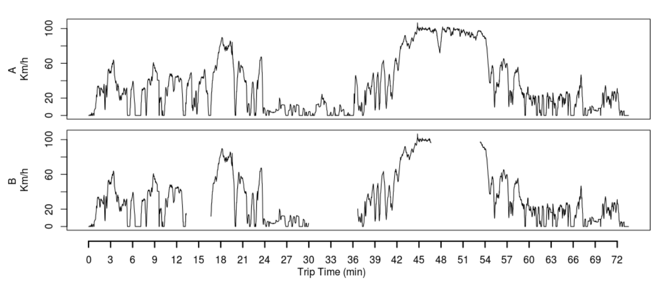

Fig. 3. Comparison Between GPS Speed and Incomplete GPS Speed Data

F. Ambiguity

Ambiguity in data sets is a manifestation of its imprecision and it happens when two occurrences in the data set are assumed to be precise and exact, however they differ from each other.

Different sensors can be considered as vehicle speed by ECU and GPS. In this case, the ambiguity manifests when both sensors present the same data to the same observation of environment. In the Figure 4, we show a histogram of absolute difference between vehicle speed and GPS speed. The major frequency of this difference is concentrated in 0 (zero), implying that both sensors collected the same speed. Furthermore, the values different to 0 implies that vehicle speed shows the current speed and GPS speed a different or conflicted value.

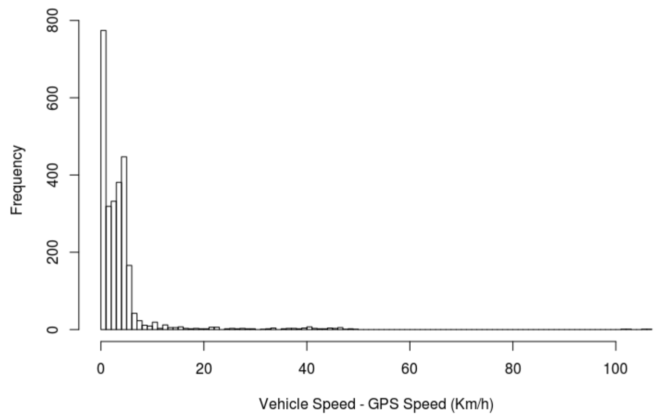

Fig. 4. Difference Between Vehicle Speed and GPS Speed

G. Uncertainty

Data collected from sensors or external sources are associated with a confidence degree. Whenever this confidence is lower than 100%, the data is considered uncertain. Solutions to this problem include statistical inference and belief functions.

In the case of sensors, the uncertainty is always present, in other words, it is inherent a property of any sensor. Even though sensor data are collected directly from the vehicle by OBD, these data are not considered an absolute true to provide a low uncertainty degree.

H. Correlation

Data correlation is problematic in data fusion, since it can either enhance or attenuate some aspects due data incest. Data incest is a situation when correlated data is fed multiple times to the data fusion system, multiplying its importance on the final result.

We perform the Pearson Product Moment Correlation (PPMC), between all sensors readings in the data collected during a trip of one vehicle, as shows in Figure 5. The matrix address the explicit values of the correlation between a par of sensors data collected from the vehicle and smartphone. We considered a high correlation values between 0.5 to 1.0

or -0.5 to -1.0, medium correlation between 0.3 to 0.5 or -0.3 to -0.5, low correlation between 0.1 to 0.3 or -0.1 to -0.3 and no correlation when 0.

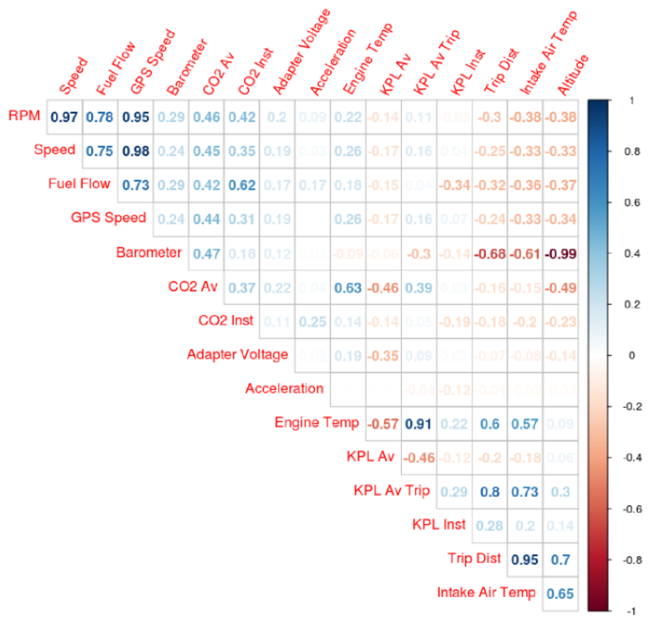

Fig. 5. Correlation Between Sensors Data in a Vehicle

Looking for high correlations, it is possible to see that revolutions per minute (RPM), Speed and GPS Speed represent the vehicle motion. So that, these data can be reduced to only one variable as Speed, for instance. However, there is a less explicit yet important relationship, like RPM and speed, which is governed by the transmission system of the car. Other possible reduction can be done in relation between altitude and the atmospheric pressure, labeled as "Barometer". It is physically proven that the atmospheric pressure is inversely proportional to the altitude. Thus, these two variable can be explained using only one.

I. Disorder

When processing continuous data sources, sometimes measurements arrive out of their order and raise a natural question: what to do with this piece? A simplistic way of treating disordered data is to simply discard it, however, this tactic would ignore the contributions of the discarded piece. A more costly solution is to store all received data and reorder the entire set once an out of order observation arises.

This problems is not common in our scenarios, because the process to data collect is synchronous and it is started by the smartphone. Other point is that the communication protocol, deals to this problem.

J. Disparateness

Vehicular sensor data is inherently disparate, since there are sensors that assess different aspects in different units and scales. Using large quantities of diverse data allows the extraction of contextual information unable to be captured by physical sensors.

As mention before, the vehicular sensor data is inherently disparate. In the vehicle, there are since sensors to measure the engine temperature until sensor to measure the fuel level. For instance, in the Figure 6, it shows a dissimilarity between two sensors as revolution per minute (RPM) and carbon dioxide emission (CO_2). It may be possible to study the behavior of these two variables, but they remain disparate.

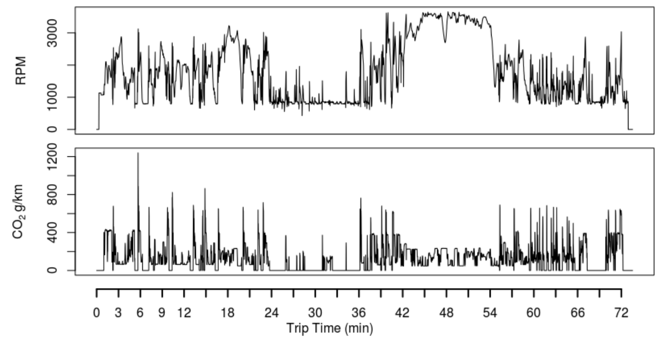

Fig. 6. Disparateness Between Revolution per Minute and Carbon Dioxide

V. Conclusion

With the constant growth of global population, urban mobility aspects and problems have become more challenging. Given the need of people to make their commutes quickly and safely in big cities, their current traffic infrastructures, and the elevated costs of restructuring it, a new approach to handle these issues is needed. Current information technologies and systems are capable of acquiring and processing massive volumes of data and outputting results with minimal delays, which makes them suitable for managing and planning new intelligent transportation systems for major cities.

ITS can be boosted by take in account heterogeneous data collected from several sources as much as possible. However, in general, the data comes with some issues (*i.e.*: imperfection, correlation, inconsistencies, among others) making difficult heterogeneous data fusion process. In this work, we conducted an exploratory analysis over real vehicle data to show, for each listed data issues, which of them were found in our data set. Indeed, we found out several issues in our data implying that they must be treated before fusion process.

It also can guide beginner researchers to better understand the data, mainly in vehicular context, and some problems they possibly have to deal.

As future extensions of this work, we will apply heterogeneous data fusion techniques to enhance, for example vehicular mobility traces by adding contextual information such as traffic conditions and driver behavior. In addition, we intend to explore data from other sources than intra-vehicle, thus we can expand possibilities of data analysis.

References

[1] E. F. Nakamura, A. a. F. Loureiro, and A. C. Frery, "Information fusion for wireless sensor networks," *ACM Computing Surveys*, vol. 39, no. 3, pp. 9–es, 2007.

[2] S. B. Ayed, H. Trichili, and A. M. Alimi, "Data fusion architectures: A survey and comparison," in *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*. IEEE, 2015, pp. 277–282.

[3] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28–44, jan 2013.

[4] N.-E. E. Faouzi and L. A. Klein, "Data Fusion for ITS: Techniques and Research Needs," *Transportation Research Procedia*, vol. 15, pp. 495–512, 2016.

[5] INRIX, "Inrix 2015 traffic scorecard," jul 2016. [Online]. Available: http://inrix.com/scorecard/

[6] ONSV, "Observatório nacional de segurança viária," jul 2016. [Online]. Available: http://www.onsv.org.br/

[7] NSC, "National safety council - our mission is safety," 2016. [Online]. Available: http://www.nsc.org/pages/home.aspx

[8] A. L. Bazzan and F. Klügl, "Introduction to intelligent systems in traffic and transportation," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 7, no. 3, pp. 1–137, 2013.

[9] S. F. Ashley Auer and S. Lockwood, "History of intelligent transportation systems," *U.S. DEPARTMENT OF TRANSPORTATION INTELLIGENT TRANSPORTATION SYSTEMS JOINT PROGRAM OFFICE*, may 2016. [Online]. Available: http://ntl.bts.gov/lib/59000/59200/59263/download1.pdf

[10] W. J. Fleming, "Overview of Automotive Sensors," *IEEE Sensors Journal*, vol. 1, no. 4, pp. 296–308, 2001.

[11] F. Qu, F. Y. Wang, and L. Yang, "Intelligent transportation spaces: Vehicles, traffic, communications, and beyond," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 136–142, nov 2010. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5621980

[12] J. Lin, S. Chen, Y. Shih, and S.-h. Chen, "A study on remote on-line diagnostic system for vehicles by integrating the technology of OBD, GPS, and 3G," *World Academy of Science, Engineering and Technology*, vol. 32, no. 8, pp. 435–441, 2009. [Online]. Available: http://waset.org/journals/waset/v32/v32-81.pdf

[13] R. R. Yager, "On the dempster-shafer framework and new combination rules," *Information Sciences*, vol. 41, no. 2, pp. 93–137, 1987.

[14] N.-E. El Faouzi, H. Leung, and A. Kurian, "Data fusion in intelligent transportation systems: Progress and challenges–a survey," *Information Fusion*, vol. 12, no. 1, pp. 4–10, 2011.

[15] J. Weng and B.-S. Lee, "Event detection in twitter." *ICWSM*, vol. 11, pp. 401–408, 2011.

[16] V. Corcoba Magaña and M. Muñoz-Organero, "WATI: Warning of Traffic Incidents for Fuel Saving," *Mobile Information Systems*, vol. 2016, pp. 1–16, 2016. [Online]. Available: http://www.hindawi.com/journals/misy/2016/3091516/

[17] R. Kosala, E. Adi *et al.*, "Harvesting real time traffic information from twitter," *Procedia Engineering*, vol. 50, pp. 1–11, 2012.