Proactive recommendations for Intelligent Mobility

An approach based on real-time big data processing

Paulo Figueiras* — Ruben Costa* — Guilherme Guerreiro* — Luka Bradesko — Nenad Stojanovic*** — Panagiotis Georgakis**** — Efthimios Bothos***** — Babis Magoutas***** — Akrivi Kioussi******

**CTS, Uninova, Dep.ª de Eng.ª Electrotécnica, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal*
paf@uninova.pt
rddc@uninova.pt
g.guerreiro@campus.fct.unl.pt

***Institut Jozef Stefan, Slovenia*
luka.bradesko@ijs.si

****Nissatech Innovation Centre, Serbia*
nenad.stojanovic@nissatech.com

*****Sustainable Transport Research Group, University of Wolverhampton, UK*
P.Georgakis@wlv.ac.uk

******National Technical University of Athens, Greece*
mpthim@mail.ntua.gr
elbabmag@mail.ntua.gr

*******Intrasoft International, Belgium*
Akrivi.Kiousi@intrasoft-intl.com

ABSTRACT. *The transportation sector is undergoing a considerable transformation as it enters a new landscape where connectivity is seamless and mobility options as well as related business models are increasing. Besides technological advancements, which have revolutionized transportation supply with services such as, real-time traveller information, collective travelling, provision of way findings and navigation services regarding multimodality of trips, the need for more efficient transportation systems emerges due to the* *current vast development in urbanization. The work presented here, presents the approach to be followed within OPTIMUM, a project that operates in an environment of ubiquitous connectivity throughout the transportation system and its surroundings that continuously provides data on the present state and emerging situations.*

KEYWORDS: *Intelligent Transport Systems, Real-time big data processing, Predictive analytics, Interoperability.*

1. Introduction

Intelligent Transportation Systems (ITS) are the means to mitigate problems emerging from the complex urban landscape including excessive CO2 emissions, high levels of congestion, increased accident risks and reduced quality of life. The emerging area of ‘digital-age transportation systems’ aims at tackling the needs of different individuals while building on massively networked systems and applying user centered approaches masking the underlying complexity. In addition, the technological developments that underpin the operation of current transportation services, leads to the generation of huge amount of available data with short update rates. This growth in data production is being driven by individuals and their increased use of media, novel types of in-vehicle and infrastructure sensors with enhanced communication capabilities, application of modern information and communication technologies with the proliferation of internet connected devices and systems. Data sets grow in size because they are being gathered increasingly by ubiquitous information sensing mobile devices, aerial sensory technologies (remote sensing), software logs, cameras, microphones, radio-frequency identification readers, and wireless sensor networks, together with machine-generated and unstructured data (e.g. photos, videos, social media feeds). This means that ITS applications have to sense huge amounts of data, process them and infer and communicate usable information such that efficient decisions and actions can be made. But for true and real change in our life, novel transport systems have to be able to foresee situations in near real time, and provide the means for proactive decisions which in turn will deter problems before they even emerge. Our vision is to provide the required interoperability, adaptability and dynamicity in systems and services for a proactive and problem-free transportation system.

OPTIMUM project (OPTIMUM consortium, 2015) operates in an environment of ubiquitous connectivity throughout the transportation system and its surroundings that continuously provides data on the present state and emerging situations. Examples include traffic and in-vehicle sensors, positioning information (e.g. GNSS data), occupancy of public transportation, crowd data sourcing through social networks and availability of modalities such as shared bicycles and cars. Sensors for vehicle-to-vehicle and vehicle-to-infrastructure information exchange offer context-aware pervasive configurations that can infer situations on transport networks. Environmental and weather sensors may provide insights on the status of the ambient environment and its impact on the utilization of the transport networks. Another important source of information refers to social sensors, i.e. citizens interacting in social media (e.g. twitter and Facebook) openly offering opinions and observations. On the other hand, user centered transportation information needs to be able to accommodate the individual. It needs to provide choices for the user in a personalized and understandable manner.

OPTIMUM’s main objective, is the establishment of a largely scalable, distributed architecture for the management and processing of multisource big-data, enabling continuous monitoring of the transportation systems needs and providing data-driven mobility services based on proactive decisions and actions in an (semi-) automatic way.

This paper is structured as follows: Section 2 presents the related work. The methods used for data fusion and harmonization are briefly explained in section 3. Finally, section 4 presents our preliminary conclusions.

2. Related Work

In order to cope with challenges and objectives identifies, OPTIMUM aims to develop a big-data fusion platform for integrating heterogeneous, static, real-time and dynamic streams of transportation data offered by public authorities, businesses and organizations as well as social media for proactive suggestions and management of transportation networks. The objective is to enable data fusion and harmonization tools supported by online stream analytics for the generation of a transport data repository that can underpin the operation of novel mobility services.

The diversity of transportation services and interoperability is another important challenge in such environment, since it is necessary to cope with a large diversity of mobility related data / services developed using different technologies. The adoption of common standards and mechanisms, the strong contribution to existing standards as well as a vendor-independent philosophy are expected to lower this barrier, which nevertheless cannot be underestimated.

2.1. *Big Data Fusion*

The term big data is used for extremely large and complex data sets, which cannot be handled properly by traditional approaches and tools. Big Data represents the information assets characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value (De Mauro, Greco, & Grimaldi, 2015). As the definition behind, the three aspects most considered to define big data in the literature are: (i) Volume, from the ever augmenting data collected; (ii) Velocity, from the growth on data acquisition; (iii) Variety, from heterogeneity of data formats and protocols used.

Other aspects that must be considered are the issues in transporting data in a big data context, such as the many inconsistencies of the data itself (from gathering values from broken road sensors, for example), outdated data, the speed variation of the internet connection, or others. Leading to a low veracity ratio, since the quality of data is constantly changing, even from the same source, and big complexity from the difficulties to fuse large amounts of data from many sources at the same time (Volume to Variety ratio).

Data needs to be gathered, clean, transform and stored efficiently, minimizing big data characteristics, in particular the exponential increase in complexity of computing or data mining, as the space required for storing. Only then, is possible to extract the value of a large amount of information. So, data needs to complete the so called big data lifecycle. ETL, introduced by Data Warehousing, stands for Extract-Transform-Load and represents the process in which data is loaded from a source to a unified data repository. ETL supporters have been extending their solutions to provide big data extraction, transformation and loading between big data platforms and traditional data management platforms, describing ETL now has “Big ETL” (Caserta & Cordo, 2015).

2.2. *Big Data Harmonization*

With respect to data harmonization, OPTIMUM adopts a comprehensive observation of the transport ecosystem, by designing and developing a smart sensing system able to cope with a huge amount of heterogeneous data in real-time. Historical and real-time data from infrastructure-based ITS, on-board units, transport operators, air-quality and weather stations, together with traveler behavioral information and crowdsourcing data will form the heterogeneous inputs to OPTIMUM platform.

In order to manage the several heterogeneous data sources, OPTIMUM adopts a standard based compliance. Therefore, DATEX is being currently addressed by the project. DATEX (EIP/EIP+ Project, 2014) standard, was first published in the end of 2006 and acknowledged in 2011 by the European Technical Specification Institute (ETSI) (European Telecommunications Standards Institute, 2015) for modelling and exchanging ITS related information, being a European standard for ITS since then. From the beginning it has been developed to provide a way to standardize information covering the communication between traffic centers, service providers, traffic operators or media partners.

Since the first release many aspects have been improved (current version is 2.3) and now offers many other features. Is at this time developed and maintained by the EasyWay project (Easyway, 2011) and supported by the European Commission. Some of the main uses are: (i) Routing/ rerouting using traffic management; (ii) Linking traffic management and traffic information systems; (iii) multi-modal information systems; (iv) information exchange between cars or between cars and traffic infrastructure systems.

3. Data Fusion and Harmonization methods

This section describes the services developed for data cleaning, fusion and harmonization. As previously stated, it is fundamental the usage of big data tools and methodologies for data management, as well as data cleaning and harmonization, as the ETL methodology, which ensures processes for extract raw data from a source, clean the data, perform transformations and load into a target. In Figure 1, is shown an overview of the methodology used for data fusion and harmonization.

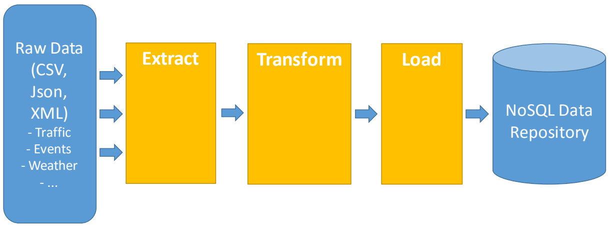

Figure 1. *ETL methodology for data fusion*

For storing, a NoSQL (Not Only SQL) base repository will be adopted as data storage, such as MongoDB, which presents advantages when handling distributed storage capacity and optimized processing for large amounts of data, which helps in scalability issues. In this non-relational database, traditional highly structured tables are replaced by collections of documents that doesn't have a specific format record set. NoSQL databases are increasingly used in big data and real-time web applications.

In terms of data harmonization, the DATEX II standard was chosen for harmonizing traffic related data. DATEX II is a complete standard for ITS related information, most generally used today and very adaptable, being able to assemble the main required types of data. All information is mapped to XML, so it is easy to transport and retrieve, regardless of the technologies used. Table 1, highlights some transformations performed at the data level, in order to be DATEX-II compliant.

Table 1. *heterogeneous sources to DATEX-II transformations*

Structure Name	Fields	Standard Correspondance	Data Type
roadSensorValue	sensorId	**datex:**SiteMeasurements (measurementSiteReference)	String
	dateTime	**datex:**DateTimeValue (dateTime)	Timestamp
	totalPerHour	**datex:**TrafficFlow (vehicleFlowValue)?	Integer
	occupancy	**datex:**TrafficConcentration (occupancy)	Integer
	Flow	**datex:**TrafficFlow(vehicleFlowValue)	Float/Integer

4. Conclusions and Future Work

The OPTIMUM project is still in its early stages and preliminary results achieved do not address the final conclusions of the project, but prepare the platform in terms of data harmonization and performance required for higher level functionalities. The aim of the work presented, is to provide a set of data operators used to implement higher level tasks like data analytics, forecasting, process optimization, complex event detection and decision making.

We strongly believe that OPTIMUM, offers a novel big data-driven solution, which will lead to seamless transportation services for the movement of people and goods. The fusion of transportation data and their transformation into actionable information, meets the mobility needs of passengers and ensures a wider choice of transportation services. Having access to proper transportation related information allows European citizens to make better use of the existing infrastructure when travelling, whereas through information personalization and persuasive mechanisms a shift to more environmentally friendly modes of transport can be achieved.

5. Acknowledgments

The authors acknowledge the European Commission for its support and partial funding and the partners of the research project: H2020–644715 OPTIMUM.

6. References

Caserta, J., & Cordo, E., “Big ETL: The Next 'Big' Thing”. Retrieved from DataInformed: http://data-informed.com/big-etl-next-big-thing/, February 9, 2015.

De Mauro, A., Greco, M., & Grimaldi, M., “What is big data? A consensual definition and a review of key research topics”. *proceedings of the 4th International Conference on Integrated Information*, 2015, Madrid, pp. 97-104, AIP.

Easyway., DATEX II –The standard for ITS on European Roads. Retrieved from http://www.datex2.eu/, 2011

EIP/EIP+ Project., Datex Easyway. Retrieved from DATEX support web site: http://www.datex2.eu/, 2014.

European Telecommunications Standards Institute. Retrieved from ETSI: http://www.etsi.org/, 2015

OPTIMUM consortium. OPTIMUM Project. Retrieved from http://optimumproject.eu/, 1 October 2015.