Cloud-based Data Analytics on Human Factor Measurement to Improve Safer Transport

Mobyen Uddin Ahmed *[1], Shahina Begum[1], Carlos Alberto Catalina[2], Lior Limonad[3], Bertil Hök[4], Gianluca Di Flumeri[5]

[1]School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden
[2]ITCL Polígono Industrial Villalonquéjar c/López Bravo, 70. 09001 BURGOS, Spain.
[3]Smart Wearable and IoT Solutions, IBM Research – Haifa, Israil
[4]Hök Instrument AB, Flottiljgatan 49, SE-72131 Västerås, Sweden
[5]Cognitive States in Operative Environment, BrainSigns, Via Sesto Celere, 7/C Rome, Italy

{mobyen.ahmed,shahina.begum}@mdh.se,{carlos.catalina}@itcl.es,{LIORLI}@il.ibm.com,{bertil.hok}@hokinstrument.com,{gianluca.diflumeri}@brainsigns.com

Abstract. Improving safer transport includes individual and collective behavioural aspects and their interaction. A system that can monitor and evaluate the human cognitive and physical capacities based on human factor measurement is often beneficial to improve safety in driving condition. However, analysis and evaluation of human factor measurement i.e. demographics, behaviour and physiology in real-time is challenging. This paper presents a methodology for cloud-based data analysis, categorization and metrics correlation in real-time through a H2020 project called SimuSafe. Initial implementation of this methodology shows a step-by-step approach which can handle huge amount of data with variation and verity in the cloud.

Keywords: SimuSafe, safer transport, data-analysis, big data, human factor

1 Introduction

As it can be found in [1-2], there are around 90% of road-traffic crashes caused by driver error (i.e. as inattention, loss of vigilance, mental under/overload) and unsafe behavior (i.e. inadequate training or lack of experience). Improving road safety includes understanding the individual, collective and interaction behaviour of drivers and pedestrians. A system that can monitor and evaluate the human cognitive and physical capacities based on human factor measurements is often beneficial to improve safety in driving condition and, more in general, in the whole transportation domain [14]. Due to increased data volume, real time data acquisition and heterogeneous sources data analytics i.e., data processing, analyzing and visualizing is becoming a challenging task. Several authors have focused on data analytics platform based on ongoing challenges [3, 4]. Most of these challenges are about real-time processing, handling of massive

* Mobyen Uddin Ahmed, Mälardalen University, SE-72123 Västerås, Sweden, mobyen.ahmed@mdh.se.

data, storage capacity, processing speed and so on. Companies like Google and Amazon have been trying to overcome these challenges using Hadoop or similar exiting data systems [5, 6].

This paper presents a methodology for cloud-based data analysis, categorization and metrics correlation in real-time through a H2020 project called SimuSafe[1]. The goal of SimuSafe is to identify behavioural models of drivers and pedestrians in a real traffic environment, implemented within traffic simulators with controllable settings, by applying artificial intelligence, virtual reality and data science methodologies. The proposed approach presented here shows the possibility to handle, process and analyze large amount of demographics, behavioural and physiological data with various variations coming both in offline and in real-time. Here, the proposed approach is implemented in IBM Bluemix cloud platform where the data analysis will be conducted in three Phases: (1) Information fusion and data abstraction, (2) Data mining and knowledge discovery and (3) Learning, reasoning and model creation.

2 System Overview

The data analytics will comprise with a data storage infrastructure to gather all relevant data (actor model states, user, cognitive and behavioural assessment data and annotations). This infrastructure will be integrated in IBM Bluemix cloud platform, further processing will be performed in the Data Analysis Server as presented in Figure 1.

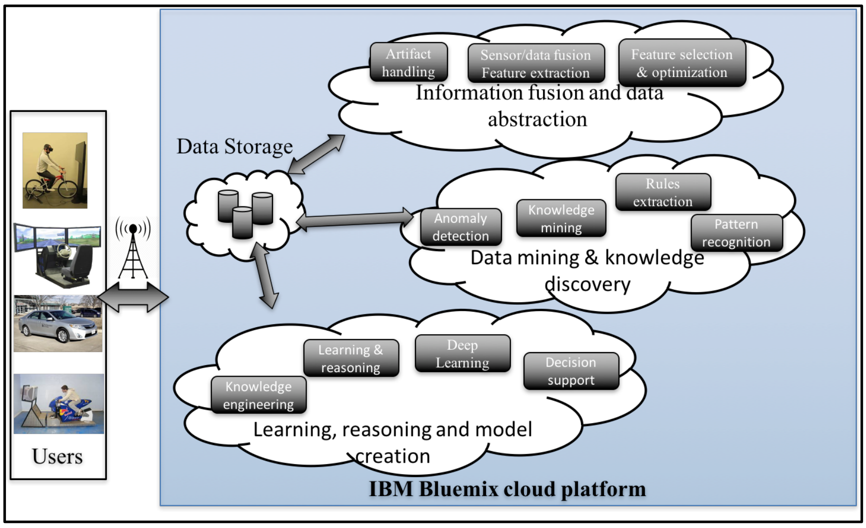

Fig. 1. Cloud-based data analytics approach

1 www.simusafe.eu

IBM Streaming and Predictive Analytics[2] from the platform will be employed to pre-filtering raw data sent to the Data Analysis server, for the real-time identification of events of interest and characteristic data patterns to be translated into components of the Actor Model. At the end of the analysis cycle, the correlation of relevant descriptors, patterns and states from the Actor model will be determined quantitatively, so responsible factors can be translated to high/low risk metric indices and used for calculation. Additionally, this data analysis will determine the cross-correlation and interdependency of data descriptor within the actor model. Tracing the effects and relations between the actor model components is essential since not all sensors are presented in Naturalistic Driving conditions (i.e. biometric data such as EEG, BVP, etc). This approach will allow the computation of the actor model and risk metrics with a sensor subset, effectively removing the need of tests with a high degree of sensorization in later stages.

3 Materials and Method

The project is going to use three research cycles that will allow to reproduce and refine the model and metric in each cycle based on the measurements collected both considering Naturalistic Driving (ND), Controlled Environment (CE) and Simulation Driving Tests (SD). There will be two test groups preforming the test, constituted by pedestrian and both two-wheels and car drivers, with young adults (18-24 years) and elderlies (50-70 years). Total test subjects are about 458, in 1st cycle 90, 2nd cycle 42 and 3rd cycle 326 across the Europe (i.e. Spain, Sweden, France, Italy, UK, Poland). The sensor measurements will be collected derived from the Human, Vehicle and Environment components, however, in this paper we only consider Human factors. Human factor measurement will be collected from the subjects and is organized in three classes: Demographics, Behavioural and Physiological data, as summarized in Table 1.

Table 1. Summary of measurements parameters related to human factor.

Demographics	Behavioural	Physiological
• Gender • Age • level of driving experience • car ownership • frequency of driving	• Propensity for aggressive driving • Sleep Hygiene • Psychological stressors • Driving Style • Incident/Violation Occurrence • Situation Awareness • Stress	• Electrooculogram (EOG) • Electroencephalogram (EEG) • Electrocardiogram (ECG) • Electromyography (EMG) • Galvanic Skin Response (GSR) • Blood Volume Pulse (BVP) • Heart Rate Variability (HRV) • Skin Temperature • Eye Tracking • Breath Alcohol

The proposed cloud-based data analysis on human factor measurement to obtain required performance works in three Phase: **Phase 1:** Information fusion and data abstraction, **Phase 2:** Data mining and knowledge discovery and **Phase 3:** Learning, reasoning and model creation.

---

2 https://console.bluemix.net/catalog/?category=data

In *Phase 1*, the data pre-processing will be performed based on a combination of statistical, machine leaning and signal processing methods and techniques. Here, a robust and scalable data cleaning strategy will be established based on domain-specific knowledge, which will include sub-processes like cleaning, filtering, sampling or/and normalization. Our previous work on data pre-processing [9] using both structured and unstructured data will serve as a basis for online automated data cleaning. Traditional feature extraction methods [7] will be adapted to handle scalability issues in the domain. In our proposed work, we will devise novel strategies to fuse data at feature level and as well as at data level considering a defined fusion mechanism [8].

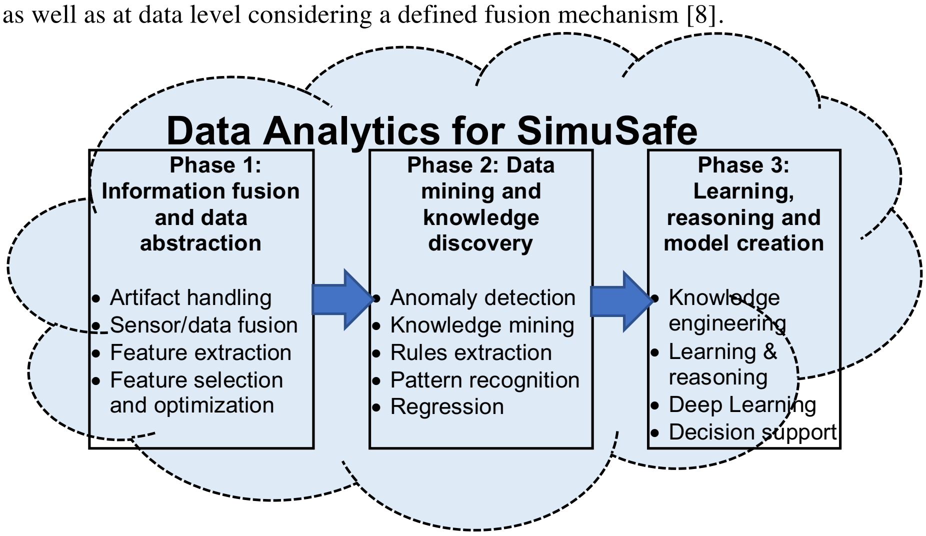

Fig. 2. Phases of the proposed data analytics approach in SimuSafe

In *Phase 2*, a combination of potential sequences in the learning and search procedure will be investigated. The similarity assessment in the time series will be done by measuring the distance between probability distributions in the time series data mining [10]. A combination of statistical model and fuzzy modeling algorithm will be applied to automatic addition/deletion of rules, as well as adjustment of the membership functions. A continuous learning procedure will be developed so as to keep the model constantly updated [12]. In addition, new mining methods to support the discovery of knowledge [12] will be developed.

In *Phase 3*, adaptation of dynamic knowledge representation approaches will be achieved by combining different artificial intelligence [12] [13] methods. This has a connection with Phase 2 as the data driven knowledge, rules and patterns will be considered as input. To provide decision support a hybrid approach will be applied utilizing different traditional machine learning algorithms, such as case-based reasoning, and clustering [11].

4 Summary

This paper proposed an approach for cloud-based data analysis, categorization and metrics correlation in real-time through a H2020 project called SimuSafe. The goal of the proposed approach shows the possibility to handle, process and analyze the Demographics, Behavioural and Physiological data in Big data contest. IBM Bluemix cloud platform is used with three parallel nodes where analytics phases are implemented. The phases are: 1) Information fusion and data abstraction, (2) Data mining and knowledge discovery and (3) Learning, reasoning and model creation. SimuSafe project is in its initial phase started in June 2017, several challenging works is ongoing.

Acknowledgments. The authors would like to acknowledge the project SimuSafe, the project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 723386.

References

1. Elander, J., West, R, French D, "Behavioural correlates of individual differences in road-traffic crash risk: An examination of methods and findings", *Psychological bulletin* 113.2 (1993): 279
2. Feyer, A.M., Williamson, A. & Friswell, R., “Balancing work and rest to combat driver fatigue: an investigation of two-up driving in Australia”. Accident Analysis and Prevention 1997 Jul, 29,541–553.
3. Katal, M. Wazid, and R. H. Goudar, "Big data: Issues, challenges, tools and Good practices," in *Contemporary Computing (IC3), 2013 Sixth International Conference on*, 2013, pp. 404-409.
4. Zhang D., "Inconsistencies in big data," in *Cognitive Informatics & Cognitive Computing (ICCI*CC), 2013 12th IEEE International Conference on*, 2013, pp. 61-67.
5. Rathore M. M., Ahmad A., Paul A., and Daniel A., "Hadoop based real-time Big Data Architecture for remote sensing Earth Observatory System," in *2015 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2015, pp. 1-7.
6. Xhafa F., Naranjo V., and Caball S., "Processing and Analytics of Big Data Streams with Yahoo! S4," in *2015 IEEE 29th International Conference on Advanced Information Networking and Applications*, 2015, pp. 263-270.
7. Begum S., Barua S., Filla R., Ahmed M.U., Classification of physiological signals for wheel loader operators using Multi-scale Entropy analysis and case-based reasoning. Expert Systems with Applications 41, pp. 295-305, 2014.
8. Begum S., Barua S., Ahmed M.U., Physiological Sensor Signals Classification for Healthcare Using Sensor Data Fusion and Case-Based Reasoning. Sensors 14, 2014.
9. Barua S., Begum S., Ahmed M.U., Clustering Based Approach for Automated EEG Artifacts Handling. 13th Scandinavian Conference on Artificial Intelligence (SCAI’15), 2015.
10. Fu. T.-C. A review on time series data mining, Engineering Applications of Artificial Intelligence, Vol. 24, pp. 164-181, 2011.

11. Ahmed M.U., Funk. P., A Computer Aided System for Post-operative Pain Treatment Combining Knowledge Discovery and Case-Based Reasoning. 20th Intl. Conf. on Case-Based Reasoning, ICCBR'12, 2012.
12. Banaee H., Ahmed M.U., and Loutfi A., Data mining for wearable sensors in health monitoring systems: a review of recent trends and challenges. Sensors, 13:17472–17500, 2013.
13. Ahmed M.U., Banaee H., Loutfi A., Health monitoring for elderly: An application using case-based reasoning and cluster analysis. ISRN Artificial Intelligence, 2013.
14. Arico, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., Colosimo, A., & Babiloni, F. (2017). Passive BCI in Operational Environments: Insights, Recent Advances and Future trends. IEEE Transactions on Biomedical Engineering.