TOWARDS SITUATION-AWARENESS AND UBIQUITOUS DATA MINING FOR ROAD SAFETY: RATIONALE AND ARCHITECTURE FOR A COMPELLING APPLICATION

Shonali Krishnaswamy[1], Seng Wai Loke[1], Andry Rakotonirainy[2], Osnat Horovitz[1] and Mohamed Medhat Gaber[1]

[1]Faculty of Information Technology
Monash University
{Shonali.Krishnaswamy, Seng.Loke, Mohamed.Medhat.Gaber}@infotech.monash.edu.au,
osnat.horovitz@gmail.com

[2]Centre for Accident Research and Road Safety Queensland
Queensland University of Technology
r.andry@qut.edu.au

Abstract - **Road crashes cost Australia $15 billion a year and 95% of these are attributed to drivers' errors. Risk assessment is at the core of the road safety problem. This paper presents an Advanced Driving Assistance System (ADAS), called SAWUR, that analyses situational driver behaviour and proposes real-time countermeasures to minimise fatalities/casualties. The system is based on Ubiquitous Data Mining (UDM) concepts. It fuses and analyses different types of information from crash data and physiological sensors to diagnose driving risks in real-time. The novelty of our approach consists of augmenting the diagnosis through UDM with associated countermeasures based on a context awareness mechanism. In other words, our system diagnoses and chooses a countermeasure by taking into account the contextual situation of the driver and the road conditions. The types of context we exploit include vehicle dynamics, drivers' physiological condition, driver's profile and environmental conditions. The rationale for exploiting contextual information is to increase the accuracy of the diagnosis (90%) and to reduce false alarm rates (below 1%). The ultimate goal is to decrease driver's exposure to risks.**

Keywords – **Ubiquitous Data Mining, Situation Awareness, and Road Safety.**

1. INTRODUCTION

Road crashes cost Australia $15 billion a year [1]. Information Communication Technology offers new safety solutions. It is estimated that Intelligent Transport Systems (ITS) could reduce fatalities and injuries by 40% across the Organisation for Economic Co-Operation and Development (OECD) and thereby saving over USD 270 billion per year [12]. Existing ITS initiatives in the area of Advanced Driving Assistance Systems (ADAS) address generic road safety problems such as rear-end collision avoidance, road departure collision avoidance and vision enhancement. These systems analyse data from various sensors to assist drivers. They improve driving performance by analysing the current situation and assessing the probability of crashes. These promising systems could prevent imminent crashes with certain accuracy based on car trajectory. However these systems do not explain why a crash is imminent. Contextual information about the driver could help to explain why a crash is imminent and improve the accuracy of crash prediction. For example existing ADAS could be augmented with information about driver's physiological state (fatigue, stress) or about the locations where crashes are occurring at high rate (black spots) to improve the accuracy of the prediction.

This paper presents the architecture of a Ubiquitous Data Mining (UDM) based on-board system, capable of analysing various type of contextual information in real-time, anywhere, anytime and with limited computational power. We call our system SAWUR (**S**ituation-**A**wareness **W**ith **U**biquitous data mining for **R**oad safety, pronounced "saviour"). SAWUR comprehensively incorporates and analyses contextual information related to driver behaviour, driver physiological and psychological profile, car dynamics and environmental information in a real time and in ubiquitous condition. This multidisciplinary approach integrates recent models of data mining, context-awareness computing, physiological metrics, ubiquitous computing, driver distraction models, risk perception and road safety. It yields a new understanding of driver behaviour and countermeasures in risk situations.

Section 2 is an overview of UDM. Section 3 shows how UDM is used to improve road safety. Section 4 discusses rationale for context awareness concepts and UDM to be integrated. Section 5 presents the design architecture of SAWUR followed by conclusions in Section 6.

2. UDM OVERVIEW

The widespread use of mobile devices with increasing computational capacity and proliferation of wireless networks is leading to the emergence of the *ubiquitous* computing paradigm that facilitates continuous access to data and information by mobile users with handheld devices.

Ubiquitous computing environments are subsequently giving rise to a new class of applications termed *Ubiquitous Data Mining* (UDM), wherein the mobile user performs intelligent analysis and monitoring of data [19, 11, 2, 10]. UDM is the process of analysing data emanating from distributed and heterogeneous sources with mobile devices or within sensor networks and is seen as the "next natural step in the world of ubiquitous computing" [7]. The techniques that are used to perform analysis typically include traditional data mining techniques that are drawn from a combination of machine learning and statistical approaches. However, these techniques have to be adapted to deal with constraints in resources, the need to perform analysis in real-time and to deal with data that is in the form of a continuous stream rather than in a database.
UDM represents the next generation of data mining systems that will support the intelligent and time-critical information needs of mobile users and will facilitate "anytime, anywhere" data mining [10, 11, 4]. The underlying focus of UDM systems is to perform computationally intensive mining/analysis techniques in mobile environments that are constrained by limited computational resources and varying network characteristics [7]. A key issue in many real UDM applications is the need to perform synthesis and knowledge integration from multiple data streams in a resource-constrained environment.

The ever-increasing computational capacity of mobile devices presents an opportunity for intelligent data analysis in applications and scenarios where the data is continuously streamed to the device and where there are temporal constraints that necessitate analysis "*anytime, anywhere"* [10, 11, 16]. Typical application scenarios are the analysis of data from sensors in moving vehicles to prevent fatal accidents through early detection by monitoring and analysis of status information [9].

It must be noted that ubiquitous data mining is not equivalent to performing traditional data mining tasks on a resource-constrained device, but addresses the unique needs of applications that require analysis of data in a time-critical and mobile context.

The typical *modus operandi* for UDM systems is for a resource-constrained device (such as a handheld device or a nominated sensor node in a network) to contain the UDM software and for this device to receive data as a stream from either sensors that continuously read environmental parameters or from external sources such as a stock exchange. The UDM module then needs to perform continuous analysis and either pass on the relevant information to a centralised component for aggregation or retain the model that is developed for local activities such as prediction and personalisation.

The opportunity and potential benefits of UDM also brings with it several challenges and research questions. This includes:

- Developing data mining techniques that can effectively analyse continuously streaming data. In this context, we note that we have developed a novel approach that incorporates adaptation of the mining process [6, 5];
- Managing user interactions and visualisation of results particularly via the limited screen real estate of small devices [11] ;
- Efficiently representing and communicating data mining models via wireless networks that are both variable and have limited bandwidth [11, 8].

3. APPLYING UDM FOR ROAD SAFETY – CURRENT STATUS AND ISSUES

The use of data mining to improve road safety can be categorised into two major approaches. The first approach concentrates on mining crash data, which includes various attributes relating to both driver and vehicle at the time of the crash [3, 17] [18]. The focus is on analysing the data for the purpose of discovering useful, and potentially actionable, information. In [18], crash data was mined to identify the driver and vehicle attributes which are the main causes for road accidents. Principal Component Analysis was used to emphasize the relationships between characteristics such as age, gender and vehicle type, to the crash variables. The paper focuses on analysing the different factors contributing to crashes, in order gain a better understanding into the causes of traffic accidents, as a means to preventing such accidents. The second major approach focuses on the area of Advanced Driving Assistance Systems (ADAS) [13, 15]. These systems concentrate on attempting to prevent specific damaging scenarios, such as vehicles rear-end collision and lane deviation. They are mostly used in Smart Cars, and they work by mining data obtained from various sensors in the car. Different data mining techniques are used in an attempt to predict a driver's moves, so that unsafe actions can be rectified, or prevented.

In [13], supervised data mining techniques, in the form of graphical models and Hidden Markov Models (HMM), have been used to create models of driving manoeuvres, such as passing, switching lanes and starting and stopping. The SmartCar applies these models to data obtained from various sensors on board the vehicle, in an attempt to predict the driver's next move. If a potentially dangerous move is predicted, the car aims to prevent the hazardous action.

Such systems, however, are limiter to very specific scenarios. They cannot identify more general dangerous driving behaviour patterns, such as driving under the influence of alcohol. Therefore, these systems cannot give advance warning, but have to wait for a critical move to be performed before they can take corrective action. Another limitation of these systems is that, on the most part, they use simulators to generate the data they then use to construct models. While simulations provide a valid strategy for collecting this information, it is no substitute for data, gathered in real-life driving conditions. According to Singh [17], if the future of Advanced Driving Assistance Systems is to be effective, real data is a necessity.

With the emergence of UDM as the next generation of data mining technology, it is evident that UDM too has the potential to play a significant role in Intelligent Transportation Systems (ITS). UDM facilitates in-vehicle analysis of sensory data received, applying classificatory approaches to event detection from sensory input and for incremental learning and model building based on sensory input in real-time.

The Vehicle Data Stream Mining System (VEDAS) [9] is a UDM system developed for real-time analysis of on-board vehicle data streams. The VEDAS system uses an on-board data stream mining and management system that focuses on deployment of the system in a mobile and resource-constrained environment. The on-board or UDM component is used to perform pre-processing of the incoming data streams to reduce the dimensionality of the data generated by the sensors using *Principal Component Analysis (PCA).* It also performs on-line unsupervised learning and implements UDM clustering algorithm. The on-board component is used to detect unusual events through this learning process. VEDAS combines the UDM component with a server component that exists at a centralised location and allows control and communication with the vehicle. It facilitation visualisation of the events in the vehicle and can be used to supplement the on-board event detection process. In transferring data between the fleet and the central server, Fourier transforms are used to compress the models that are learnt on-board prior to transfer to reduce the amount of data that is transmitted.

The VEDAS system is evidently a key milestone in the area of applying UDM principals and techniques in an ITS context for road safety. However, there are many open issues that need to be addressed in order to deliver the full promise that ubiquitous data mining has to offer for road safety:

- UDM systems need to be supplemented with context-models of on-road conditions to increase the accuracy of the response that these systems take to hazardous/unusual events.
- The use of a supervised learning or classificatory approach is one that has tremendous potential in applying UDM in a road safety situation. With such an approach, the model for identifying unusual or hazardous conditions is built off-line through traditional data mining techniques and is verified through expert users. This model is then deployed on-board the vehicle and the UDM algorithm is used to detect and classify new events as they occur based on the model available. This approach has the advantage over the clustering approach used in VEDAS, in that it can work faster in a real-time situation, as well as the model itself that is used is one that has been validated. A key issue is that in building such models to be used in the UDM based classification, it is important to collect specific data that enables detection of such events.
- The lack of available "real" data, as opposed to simulator generated data. Using simulated data to construct a model for identifying dangerous conditions has limitations, especially in relation to contextual information. Data gathered online, using unsupervised UDM techniques, will result in a model which is both more accurate and more comprehensive.
- Given the sheer number of vehicles that need to be factored in, the scalability of the data transmission issues need to be addressed. Thus, it is infeasible to assume that data from each vehicle can be transmitted in real-time and analysed. In this paper, we present a model that aims to apply UDM in a novel context – *on-board data synopsis.* We apply UDM techniques to learn in unsupervised mode on-board the vehicle – so that patterns and clusters (which capture the semantics of the data, but are compact in size) can be transmitted rather than raw data. Such a model also provides a useful live data gathering mechanism.

4. UDM AND SITUATION AWARENESS FOR ROAD SAFETY

Situation refers to the state of affairs of an entity, and in our case, would include contextual information about the driver, the vehicle, and the environment in which the car is situated. On-board UDM techniques are used to analyze readings from in-vehicle sensors in order to infer contextual information. Such contextual information are aggregated to determine what situation the driver (and the vehicle) is in, and to assess the risks involved.

Such analysis requires pre-defined models (that relate sensor readings, contextual information, and descriptions of situations) [14]. The appropriate actions (countermeasures) can then be automatically initiated to aid the driver or avoid dangers. Driver conditions (such as high alcohol levels, drowsiness, and fatigue), vehicle situations (such as nearby "dangerous" cars, lane-change, and road-departure, and intersections), and road conditions (such as vehicle traffic, wet or dry, and so on) go to make up the situation of the driver (and vehicle) and can be used to flag high risk situations and initiate countermeasures. It is a challenge to recognize these situations in the most cost-effective, timely, unobtrusive, and reliable manner and to implement and select appropriate countermeasures. It could be that the recognition of such situations results in several possible countermeasures (e.g., one for driver fatigue and another for lane-change, i.e. we have a tired driver changing lanes), of which one is most appropriate. The countermeasures themselves must not take control from the user unless necessary, and the system might need to detect that its countermeasures did work (e.g., recognizing that the driver seems drowsy and the car is about to go off the road, the system acts to alert the driver, and senses that the driver must have been alerted and the car is back on track). The situation-aware system (in the car and the supporting infrastructure) works continuously to understand the risk that the passengers of the vehicle are being exposed to and automatically acts to reduce the risks.

5.SYSTEM ARCHITECTURE OF SAWUR

The conceptual architecture of our system based on the integration of ubiquitous data mining for event/situation detection in real-time and context-awareness to guide the response/counter-measure for the detected situation is presented in Figure 1.

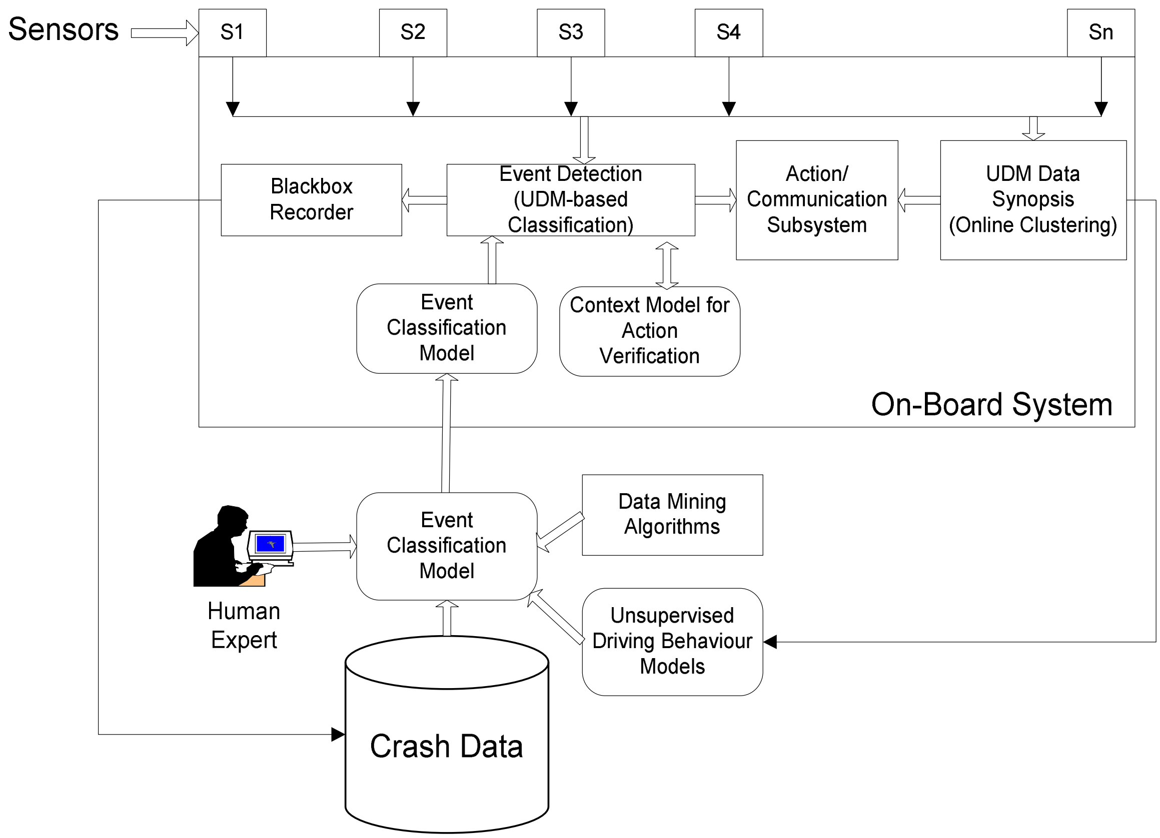

Fig. 1. Ubiquitous Data Mining and Context-Awareness for Road Safety

The system consists of a crash database that contains historical crash data. This data is mined using traditional data mining techniques to build predictive models for classifying new/unseen hazardous events. The model that is built is verified, modified and enhanced using human expertise before it is deployed onto the Ubiquitous Data Mining system on-board the vehicles. The crash database is updated from on a needs basis from event data recorded by the on-board system. This in turn leads to a cyclic approach where the predictive models are incrementally re-built and updated as new data arrives. This approach allows initial models to be refined and re-deployed in an incremental manner. Thus, the very first models may even be developed using simulation data and human expertise and refined as on-board events are recorded and used to populate the crash database with real data.

The on-board system consists of several sensors that continuously detect environmental context conditions and feed these to an UDM classificatory module. This on-board system have the ability to cope with high speed data streams and perform classification using the available predictive model in real-time given the limited computational resources that are available on-board. We have developed such UDM classification algorithms for data streams that have the ability to operate in resource-constrained environments [5].

Once a sequence of events that is classed as “potentially alarming” based on the predictive model is detected, a signal is sent to the black box component to start recording events. This is the data that is used to update the crash database. This approach addresses the need to reduce data transmission between the vehicles and the central crash database by focusing on recording data that pertains to alarming events rather than recording mundane happenings. This is obviously necessitated by the constraints of costs as well as availability associated with both hardware and communication resources. In the meantime, as a detected sequence of events indicate an escalation in potential risk levels, the context-model is used to verify and ascertain the risk levels and take remedial action as necessary. An illustrative example could be a case where a driver is on the freeway and is travelling well below the required speed. This may indicate that the driver detected as a potentially alarming situation. However, the context-model may verify that is possibly fatigued, according to the predictive model that is available. This could be there has been a block on the freeway and all the vehicles in the vicinity are also travelling at similar speeds and that there no reason for any remedial response or warning signals to be issued.

The on-board system also has an online UDM component that uses unsupervised learning techniques to create data synopsis.

By creating a data synopsis, rather than constantly recording and sending raw data, we save in both memory and communication costs. On board the vehicles, incremental clustering is performed using the LWC algorithm [5], which is a lightweight one-look incremental clustering algorithm, designed for operating in real time on resource constrained devices. On each vehicle, several different clustering models of driving behaviour are built, according to the specific spatial/temporal context. For example, there might be one clustering model for driving behaviour in the city in the morning, and another for driving behaviour in the country at night. If a model remains inactive for a certain amount of time, it is discarded, so as not to unnecessarily encumber the resource constrained device. Once a clustering model has stabilized, it is sent to a central server. In the central server, these clustering models are integrated, to provide a comprehensive general model of driving behaviour. This model is used, in combination with the crush data and the data recorded by the black box component, for constructing and updating the event classification model which is applied on-board the vehicles.

There are several practical and implementation considerations that need to be factored to realise this model. These include:

- *Data:* Multi-dimensional and multiple data streams that are generated at a rapid rate need to be analysed in real-time. This is certainly a considerable challenge given the current state of the art. Furthermore, in order to develop predictive models that can be used in classifying real-world events, it is imperative to acquire access to crash data. However, one of the challenges is that crash databases only record posterior data such as when and where the crash occurred, the driver profile and so on. We do not typically have recordings of the sequence of prior events that led to the crash. In order to obtain this type of information, there are two possible options open to us. Firstly, there is the option of using simulators to build such data. The second option is more innovative and proposes using UDM components in conjunction with black-box recorders to obtain this data. We are currently conceptualising and developing this component.
- *Analysis:* While we have developed light-weight data analysis algorithms that adapt the rate of functioning according to available computational resources [6], these algorithms need to be modified to deal with integration from multiple streams.
- *Communication:* The transfer of models that are built and the frequency of this transfer needs to be determined to minimise the data transfer overhead.
- *Computational Resources:* While there are handheld devices that have large disk storage, the model that is presented is more reliant on processor and memory resources as these are the typical overheads of analysis algorithms. The ability of the model to effectively function in these constraints needs to be experimentally established.
- *Ethics and Legalities:* The ability to deploy this model would require ethics approval for monitoring. However, experimental deployment within for a fleet is planned as part of the trial and evaluation phase.

6. CONCLUSIONS AND FUTURE DIRECTIONS

The effectiveness of behaviour change as a means to improve road safety in highly motorised countries has plateau-ed therefore it is time to seek for novel and alternatives measure that exploit advances in communication technology to reduce or cope with human errors. This paper presents a new approach based on Ubiquitous Data Mining to reduce or cope with human errors by monitoring driving risks in real time. We presented the architecture of SAWUR capable of estimating risks. The type of risks that could be potentially monitored by our system include (but is not limited) to fatigue, roll over, speed and inexperience. These risks are among the factors addressed by the Australian road safety priorities. Our next step is to implement our system. Through detailed analysis of the contextualised factors contributing to crashes we will design countermeasures that will monitor risks. The last step of this project will consist of evaluating UDM in real driving conditions and pay a particular attention to human factors (distraction). This is a position paper based on our extensive experience in road safety, data mining and context awareness systems.

REFERENCES

[1.]Bureau of Transport Economics – Commonwealth Australia, 'Road Crash Costs in Australia – Report 102', 2000 http://www.bte.gov.au.

[2.]Chen, R., Sivakumar, K., Kargupta, H., 'An Approach to Online Bayesian Learning from Multiple Data Streams', Workshop on Ubiquitous Data Mining for Mobile and Distributed Environments, Held in Conjunction with Joint 12th European Conference on Machine Learning (ECML'01) and 5th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'01), September 3-7, 2001, Freiburg, Germany. Available Online: http://www.cs.umbc.edu/~hillol/pkdd2001/papers/chen.pdf

[3.] Flach, P.A., Mladenic, D., Moyle, S., Raeymaekers, S., Rauch, J., Rawles, S., Ribeiro, R., Sclep, G., Struyf, J., Todorovski, L., Torgo, L., Blockeel, H., Wettschereck, D., Wu, S., Gartner, T., Grobelnik, M., Kavsek, B., Kejkula, M., Krzywania, D., Lavrac, N. and Ljubic, P., 'On the road to knowledge: mining 21 years of UK traffic accident reports'., Data Mining and Decision Support: Aspects of Integration and Collaboration, pages 143-155. Kluwer Academic Publishers, January 2003

[4.]Grossman, R., 'Supporting the Data Mining Process with Next Generation Data Mining Systems', Enterprise Systems, August 1998, Available Online: http://www.esj.com/back_issues/toc.asp?MON=8&YR=1998

[5.] Gaber, M. M., Krishnaswamy, S. and Zaslavsky, A., 'Cost-Efficient Mining Techniques for Data Streams', in Proc. Australasian Workshop on Data Mining and Web Intelligence (DMWI2004), Dunedin, New Zealand. CRPIT, 32. Purvis, M., Ed. ACS, 2004

[6.] Gaber, M, M., Zaslavsky, A., and Krishnaswamy, S., 'A Cost-Efficient Model for Ubiquitous Data Stream Mining', the Tenth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2004), Perugia Italy, July 2004.

[7.] Hsu, J., 'Data Mining Trends and Developments: The Key Data Mining Technologies and Applications for the 21st Century', The Proceedings of the 19th Annual Conference for Information Systems Educators (ISECON 2002), ISSN: 1542-7382, Available Online: http://colton.byuh.edu/isecon/2002/224b/Hsu.pdf

[8.]Kargupta, H., and Park, B., 'Fourier Spectrum-based Approach to Aggregate and Visualize Decision Trees for Mobile Applications', Workshop on Ubiquitous Data Mining for Mobile and Distributed Environments, Held in Conjunction with Joint 12th European Conference on Machine Learning (ECML'01) and 5th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'01), September 3-7, 2001, Freiburg, Germany. Available : Online:http://www.cs.umbc.edu/~hillol/pkdd2001/papers/kargupta.pdf

[9.] H. Kargupta, R. Bhargava, K. Liu, M. Powers, P. Blair, S. Bushra, J. Dull, K. Sarkar, M. Klein, M. Vasa, and D. Handy., 'VEDAS: A Mobile and Distributed Data Stream Mining System for Real-Time Vehicle Monitoring', Proceedings of the SIAM International Data Mining Conference, Orlando, 2004

[10.]Krishnaswamy, S., Loke, S, W., and Zaslavsky, A., 'Supporting the Optimisation of Distributed Data Mining by Predicting Application Run Times**'**, Proceedings of the Fourth International Conference on Enterprise Information Systems (ICEIS 2002), April 3-6, Ciudad Real, Spain, pp. 374-381.

[11.]Kargupta, H., Park, B., Pittie, S., Liu, L., Kushraj, D., and Sarkar, K., 'MobiMine: Monitoring the Stock Market from a PDA', SIGKDD Explorations, January, Vol. 3, No. 2, 2002.

[12.]OECD: Organisation for economic co-operation and development, 'Road Safety – Impact of New Technologies', August 2003.

[13.]Oliver, N. and Pentland, A. P., Graphical Models for Driver Behavior Recognition in a SmartCar, MIT IEEE intelligent vehicles symposium, 2000.

[14.]Rakotonirainy A., 'Sustainable Context-Aware Programming for Automotive Applications', Proceeding of the Workshop on Sustainable Computing (SPC'04) in conjunction with the Second International Conference on Pervasive Computing, Vienna, Austria, April 2004.

[15.]Sharke, P., Smart Cars, mechanical engineering magazine online, September 2004, Available online: http://www.memagazine.org/backissues/mar03/features/smartcar/smartcar.html.

[16.]Soe, T.A., Krishnaswamy, S, Loke, S.W., Indrawan, M. and Sethi, D., 'AgentUDM: A Mobile Agent Based Support Infrastructure for Ubiquitous Data Mining', Proceedings of the 18th International Conference on Advanced Information Networking and Applications (AINA 2004), March 2004, Fukuoka, Japan, IEEE Press.

[17.]Singh, S., 'Identification of Driver and Vehicle Characteristics through Data Mining the Highway Crash', National Highway Traffic Safety Administration, USA, 2001.

[18.]Singh, S., 'A Sampling Strategy for Rear-End Pre-Crash Data Collection', National Highway Traffic Safety Administration, USA, 2001.

[19.]Zaki, M, J., 'Editorial: Online, Interactive and Anytime Data Mining', SIGKDD Explorations, Vol. 3, Issue 2, January 2002. Available Online: http://www.acm.org/sigs/sigkdd/explorations/issue3-2/contents.htm #Editorial