

## **Ubiquitous Data Mining for Road Safety**

Prajwala T R

### **ABSTRACT:**

The number of accidents caused every year is disastrously high 95% of these are attributed to drivers' errors. Risk assessment is at the core of the road safety problem. The system is based on ubiquitous data mining (UDM) concepts i.e a combination of association and clustering algorithms. It fuses and analyses different types of information from crash data and physiological sensors to diagnose driving risks in real time. The system diagnoses and chooses a counter measure by taking into account the contextual situation of the driver and the road conditions. The types of context include vehicle dynamics, drivers' physiological condition, driver's profile and environmental conditions. Thus it deals with proposing an innovative, intelligent system which aids the drivers take a proper decision to pre-decide their next move, hence reducing drastically the probability of many road accidents.

**Keywords:** Data mining , Ubiquitous datamining , SAWUR (Situation-Awareness With Ubiquitous data mining for Road safety), ADAS- Advanced Driving Assistance System , PCA(principal component analysis)

### **1. INTRODUCTION:**

Information Communication Technology offers new safety solutions for road safety. It is estimated that Intelligent Transport Systems (ITS) could reduce fatalities and injuries. The 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 systems could prevent imminent crashes with certain accuracy based on car trajectory. 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- Advanced

Driving Assistance System could be augmented with information about driver's physiological state or about the locations where crashes are occurring at high rate to improve the accuracy of the prediction. This system is called as SAWUR (Situation-Awareness With Ubiquitous data mining for Road safety) 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, ubiquitous computing, driver distraction models, risk perception and road safety. It yields a new understanding of driver

---

Prajwala T R currently pursuing M.Tech in PESIT college ,  
Bangalore,India  
E-mail id: prajwalatr@gmail.com

behaviour and countermeasures in risk situations.

limited computational resources and varying network characteristics.

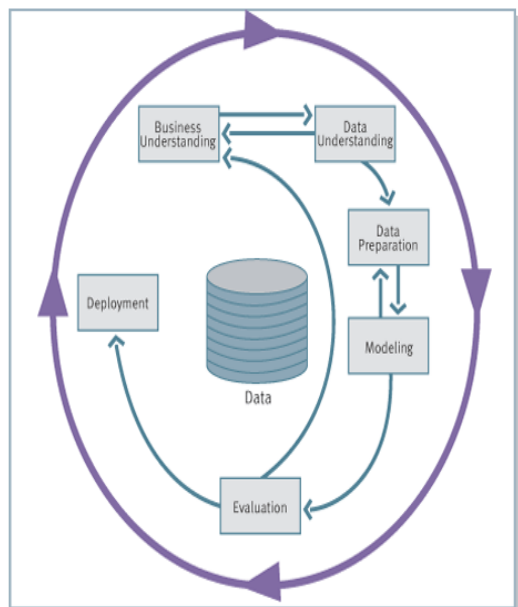


Fig1:Steps of datamining[1]

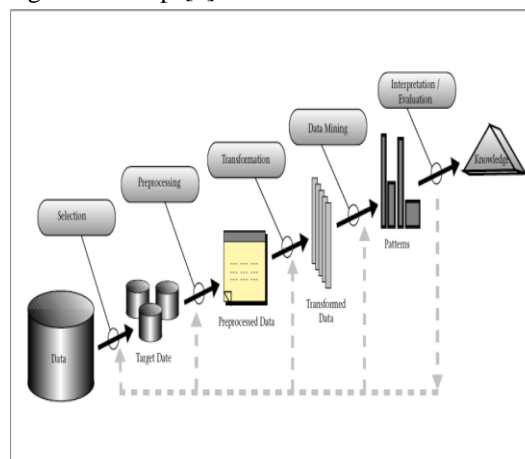
Thus the diagram shown represents the general steps of datamining. It can be applied to the road safety systems.

## 2. UDM OVERVIEW:

Ubiquitous computing environments are subsequently giving rise to a new class of applications termed *Ubiquitous Data Mining* (UDM). UDM is the process of analysing data emanating from distributed and heterogeneous sources with mobile devices or within sensor networks . 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. The underlying focus of UDM systems is to perform computationally intensive mining/analysis techniques in mobile environments that are constrained by

## UDM STEPS:

Fig2:UDM steps[2]



In general UDM process includes activities like selection, pre-processing, transformation , evaluation , knowledge.

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 . 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 .

## 3. APPLYING UDM FOR ROAD SAFETY[4]:

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. The focus is on analysing the data for the purpose of discovering useful, and potentially actionable, information. 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 second major approach focuses on the area of Advanced Driving Assistance Systems (ADAS)[4]. 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. 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. UDM has 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 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 UDM component is used to detect unusual events through this learning process.

There are many open issues that need to be addressed in order to deliver the road safety using UDM:

- 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.

This model is then deployed onboard 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 is that it can work faster in a real-time situation. We use simulated data to construct a model for identifying dangerous conditions. Data gathered online, using unsupervised UDM techniques, will result in a model which is both more accurate and more comprehensive.

#### 4. UDM AND SITUATION AWARENESS FOR ROAD SAFETY:

UDM awareness include contextual information about the driver, the vehicle, and the environment in which the car is situated. UDM techniques are used to analyze readings from in-vehicle sensors .The contextual information are aggregated to determine what situation of the driver and to assess the risks involved.The appropriate actions can then be automatically initiated to aid the driver or avoid dangers. Driver conditions[3](such as high alcohol levels, drowsiness, and fatigue), vehicle situations(like nearby "dangerous" cars, lane-change, and road-departure, and intersections), and road conditions (like 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, and reliable manner and to implement.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

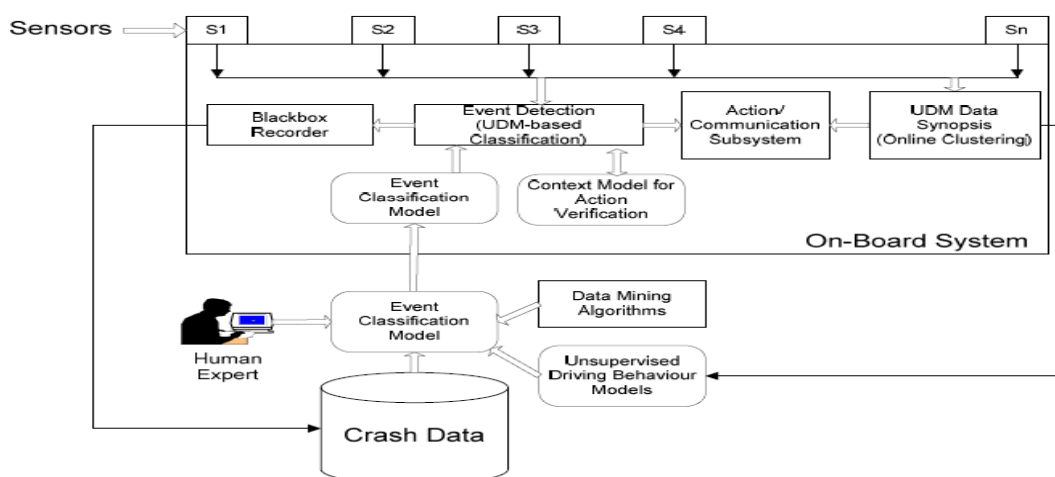


Fig.3.Ubiquitous Data Mining and Context-Awareness for Road Safety[4]

The conceptual architecture of our system based on the integration of ubiquitous data mining for vent/situation detection in real-time. 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 crash database is updated from on a needs basis from event data recorded by the on-board system. 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. 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. 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.

In the meantime, as a detected sequence of events indicate an

escalation in potential risk levels, the context-model is used to verify and as certain the risk levels and take remedial action as necessary. However, the context-model may verify that is possibly fatigued, according to the predictive model that is available. 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 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 crash data and the data recorded by the black box component, for constructing and updating the event classification model which is applied onboard the vehicles.

There are several practical and implementation considerations that need to be factored to realise this model. These include[4]:

- *Data*: Multi-dimensional and multiple data streams that are

generated at a rapid rate need to be analysed in real-time. components must be in conjunction with black-box recorders to obtain this data.

- *Analysis:* While we have developed light-weight data analysis algorithms that adapt the rate of functioning according to available computational resources, 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.

Some of the advantages and disadvantages are

### 5.1 ADVANTAGES :

- UDM technique can be applied to real time systems like the road safety system[5].
- It is a cost effective and easy to use tool for improving safety and reducing operating costs .
- Reduce injuries and prevent crashes.
- Has control on driver performance.
- Minimize risk and liability.

### 5.2 DISADVANTAGES:

- The system cannot identify more general dangerous driving behaviour patterns, such as driving under the influence of alcohol. Therefore, these systems cannot give advance warning.
- 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[5].

## 6. CONCLUSIONS:

The system is a combination of association and clustering algorithms to reduce the risks caused by road accidents. It is an approach based on Ubiquitous Data Mining to reduce or cope with human errors by monitoring driving risks in real time. The architecture of SAWUR is capable of estimating risks. The type of risks that could be potentially monitored by our system include fatigue, roll over, speed and inexperience. Thus further improvements can be made in this system to increase the road safety and counter measures that monitor the risk.

## REFERENCES:

1. Data mining techniques second edition by Michael J Berry, 2011 edition
2. Data mining methods and models by Daniel T Larose, 2012 edition
3. Gull, K.C., Intelligent Agent & Multi-Agent Systems, 2009. IAMA 2009. International Conference on, 22-24 July 2009
4. Krishnaswamy, Shonali, Loke, Seng Wai, Rakotonirainy, Andry, Horovitz, Osnat, & Gaber, Mohamed Medhat (2005) Towards situation-awareness and ubiquitous data mining for road safety: Rationale and architecture for a compelling application. In *Intelligent Vehicles and Road Infrastructure Conference*, 16-17 February 2005, Melbourne, Victoria.
5. [www.grsproadsafety.org/india](http://www.grsproadsafety.org/india)