

SURVEY OF ACCIDENT SEVERITY ESTIMATION USING DATA MINING TECHNIQUES

S. Ramya

PG Scholar: Department of CSE
SNS College of Engineering
Coimbatore, India.
jesusramya91@gmail.com

Mr.R. Senthil Kumar

Assistant Professor: Department of CSE
SNS College of Engineering
Coimbatore, India.

Abstract - As technology has been improved by many changes and inventions are made almost every day. Traumatic events come in many different ways at many different times of one's life. Recent trends on the advantage of communication technologies in case of emergency situations supported by artificial intelligence systems (AIS) which are capable to take automatic decisions. To improve the assistance time of rescue process in emergency cases with valid data for estimating severity of accident .This paper gives the analysis of accident severity using task and tools of Data Mining (DM).The estimation of accident severity based on type of accident, type of vehicle involved and person injury. An impact of our analysis includes essential features which confess to propagate estimation models and provide resources to assist the situation.

Keyword – Data Mining, Vehicle to vehicle and roadside infrastructure, Knowledge Discovery in Databases

I. INTRODUCTION

Accident is one of the several pertinent health issues currently affecting many countries. Particularly road traffic accidents are on the record for causing massive deaths in many countries. W.H.O ranks at number worldwide contributing to 1.29% of total death resulting from road accidents in the country. Road accident has claimed many lives and caused severe injuries in Australia in the recent years. World Health Organization has put road accidents at position two in the most significant cause of death for 10 to 46 year oldies.

Vehicle accidents are one of the causes of infirmity, harm and fatality. India has the highest road traffic accident rate worldwide with over 140000 deaths annually, beating even China [8]. Nearly 14 lives per hour are lost due to road accidents. Majority of accidents occur during peak hours considered an 'unsafe' or dangerous time due to complex flow pattern of vehicular traffic. Data Mining (DM) has been increasing its role popularly of substantial contribution in various applications such as fraud detection, medical, retailers and industries.

The vehicle accident is detected by integration of sensors capabilities on-board of vehicles. The minimum set of data are exchanged from sensor to estimate accident severity by selection of relevant features (type of vehicle, impact speed and status of airbag) by generating estimation models by the use of classification models in Data mining (DM) which can predict severity of new accidents and reduce assistance time of emergency services by providing recommended resources.

A. Data Mining

Data mining (also known as Knowledge Discovery in Databases - KDD) has been defined as trivial extraction of inherent, previously unknown, and possibly retrieving an information.

To discover knowledge in the form of easily comprehensible to humans some techniques are used such as machine learning, statistical and visualization.

Data mining provides an analysis of observational data sets to find unsuspected relationships and also summarize the data in novel ways that are both understandable and useful to the data owner. Data mining as a process of KDD process

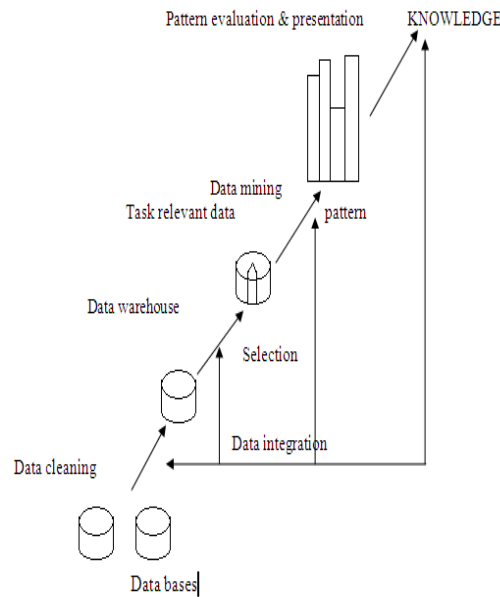


Fig. 1. Data Mining as a KDD process

B. Vehicle Accident

Accidents are generally classified as single vehicle accidents in which the vehicle is either colliding with fixed objects or with pedestrians and multiple vehicle accidents in which two or more than two vehicle can either collide or one vehicle may collide with the front vehicle at the back or may have side-swipe type collision [2]. To reduce number of road accidents, vehicular network plays major role in Intelligent Transportation System (ITS) area [3]. Road safety, fleet management and navigation such ITS applications will rely on data exchanged between the vehicle and road side infrastructure (V2I) or even directly between vehicles (V2V) [4].

II. DATA MINING TASK

Data Mining (DM) is often defined as finding hidden information in a database. Based on the task which is being performed some of mining algorithms is used to transform data to generate the desired results.

A. Classification

Classification maps data into predefined groups or classes. It is also referred as supervised learning because the classes are determined before examining the data. Classification algorithms require that the classes be defined based on data attribute values. It describes these classes by looking at characteristics of data which is already known to belong to the classes.

B. Regression

Regression is used to map a data item to a real valued prediction variable. It involves the learning of the function that does this mapping. Regression assumes that the target data fit into some known type of functions (linear, logistic,..) and determines the best function of this type that models the given data.

C. Time Series Analysis

Time series analysis is the value of an attribute is examined as it varies over time. The values which are obtained as evenly spaced time points. A time series plot is used to visualize the time series.

D. Prediction

Many real world data mining applications can be seen as predicting future data states based on past and current data. Prediction includes predicting a future state rather than current state. It refers to a type of application rather than data mining modeling.

E. Clustering

Clustering is similar to classification except that the groups are not predefined, but rather defined by the data. Clustering is unsupervised learning or segmentation. It can be of partitioning or segmenting the data into groups that might or might not be disjointed. Clustering is accomplished by determining the similarity among the data. The most similar data are grouped into clusters.

F. Summarization

Summarization maps data into subsets with associated simple descriptions. Summarization is also called characterization or generalization. It extracts or derives representative information about the database. This may accomplish that actually retrieving portions of the data.

G. Association Rules

An association rule is a model that identifies specific types of data associations. The association rules which are frequently used in the retail industry to identify items that are frequently purchased together.

H. Sequence Discovery

Sequence discovery is used to determine sequential patterns in data. The patterns are based on a time sequence of actions and are same as associations in which data are found to be related but the relationship is based on time.

Among the data mining task classification plays major role in classifying accident type severity estimation. Classification includes the process of predicting the estimation of accidents based on the given input from vehicular network. In order to predict the outcome, the algorithm processes a training set containing a set of attributes and the respective outcome, usually called goal or prediction attribute. The algorithm discovers the relationships between the attributes that would make it possible to predict the outcome.

III. METHODOLOGY

The supervised methods used are J48 Decision Trees, Bayesian network and Support Vector Machines. These methods are used for accident severity estimation.

A. Bayesian Networks

A Bayesian network is a probabilistic graphical model that represents a set of random variables and their conditional dependencies through directed acyclic graph. Bayesian network could represent the probabilistic relationships between variables such as vehicle damage and accident severity estimations.

B. J48 Decision Trees

A decision tree is a predictive machine learning model that decides the target value of a new sample based on various attribute values of the obtained data. The internal nodes of a decision tree denotes different attributes, the branch between these nodes tell us the possible values that these attributes can have in the observed samples, while the terminal nodes indicates the final value of the dependent variable. It follows simple algorithm in order to classify a new item, a decision tree has to be created based on the attribute values of the available training data. A set of items which is encountered will be identified as the attribute that discriminates the various instances most clearly. This feature includes more about the data instances so that it can be classified as it has the highest information gain. Among all possible values of this feature which has no ambiguity, for which the data instances falling within its category have the same value for the target variable, that branch is terminated and assign to it the target value that is obtained.

C. Support Vector Machines

Support Vector Machines are supervised learning methods used for classification and regression. The benefit support vector machines are that they can make use of certain kernels in order to transform the problem, so that linear classification techniques can be applied to non-linear data [7]. Apply the kernel equations which arrange the data instances within the multi-dimensional space, so that there is a hyper-plane that separates data instances of one kind from another. Support vectors instances that are either on separating planes on each side, or on the wrong side. Support vector machine is that the data has to be separated as binary. Even though data are not binary, support vector machines handles it and completes the analysis through a series of binary assessments on the data.

Among the above data mining methodology, Bayesian networks are models of the problem domain probability distribution, they can be used for computing the predictive distribution on the outcomes of possible actions. Probability theory provides a consistent calculus for uncertain inference, in which the system is always unambiguous. Given the input, all the alternative mechanisms for computing the output by using Bayesian network model produce exactly the same result.

IV. TOOLS

The tools which are used to implement data mining task as open source software.

A. Weka

Weka is a collection of machine learning algorithms for data mining tasks. The algorithm can be applied directly to a dataset and also be called using Java code [5].



Fig.2.Weka

Weka tool includes data mining task such as classification, clustering, association rules, regression, preprocessing and visualization and it is also well-suited for developing new machine learning schemes. Weka tool is open source software issued under the GNU General Public License.

B. Rapid Miner

Rapid Miner is open-source system for data mining. It is used as an application for data analysis and as a data mining engine for the integration into own products. Thousands of applications of rapid miner in more than 40 countries give their users a competitive edge.

Among the data mining tools Weka is particular tool which can assist an organization evaluate and analyze the information in more effective term. Weka tool includes the process of correlations and patterns of information. The data mining algorithm can be applied directly to dataset using Java code. Weka contains tools for data mining task such as data preprocessing, classification, regression, clustering, association rules and visualization. The Weka tool is a standard open source tool gives effective output for implementation of data mining task.

V. CONCLUSION

The communication technologies which are integrated into automotive sector offers better assistance time of rescue services at emergency situations. The vehicle accident is detected by the use of sensors capabilities by the application of Intelligent Transportation Systems (ITS) applications in which data exchanged between the vehicles (V2V) or vehicle to road side infrastructure (V2I). A prior assessment of severity of an accident is needed to adapt resources accordingly. This estimation is done based on historical data from previous accidents and also type of crash such as front crash, rear crash and side crash using data mining. The analysis includes that classification algorithms classifies the severity of accident based on the received data from sensors can noticeably increase accuracy of the system. The data mining tools such as weka and rapid miner can be used to implement data mining task.

REFERENCES

- [1] Juan de ona, Randa O.Mujali and Francisco J.Calvo(2011) Analysis of traffic accident injury severity on Spanish rural highways using Bayesian network. *Accident analysis and Prevention*, 43,402-411.
- [2] Dirección General de Tráfico (DGT). (2010). The Main Statistics of Road Accidents Spain.
- [3] J. Miller, "Vehicle-to-vehicle-to-infrastructure (V2V2I) intelligent transportation system architecture," in *Proc. IEEE Intell. Veh.Symp.*, Eindhoven, Netherlands, Jun. 2008, pp. 715-720.
- [4] F. Martinez, C.-K. Toh, J.-C. Cano, C. Calafate, and P. Manzoni, "Emergency services in future intelligent transportation systems based on vehicular communication networks," *IEEE Intell. Transp.Syst. Mag.*, vol. 2, no. 2, pp. 6-20, Oct. 2010.
- [5] M. Hall et al., "The WEKA data mining software: An update," *SIGKDD Explor.*, vol. 11, pp. 10-18, Nov. 2009.
- [6] G. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," *Mach. Learn.*, vol. 9, pp. 309-347, Oct. 1992.
- [7] J.C. Platt, *Fast Training of Support Vector Machines Using Sequential Minimal Optimization*, Cambridge, MA, USA: MIT Press, 1999, pp. , 185-208.
- [8] T. Beshah and S. Hill, "Mining road traffic accident data to improve safety: Role of road-related factors on accident severity in ethiopia," in *Proc. AAAI AI-D*, Stanford, CA, USA, Mar. 2010.
- [9] T. B. Tesema, A. Abraham, and C. Grosan, "Rule mining and classification of road accidents using adaptive regression trees," *IJSSST*, vol. 6, no. 10-11, pp. 80-94, 2005.
- [10] M. Chong, A. Abraham, and M. Paprzycki, "Traffic accident analysis using machine learning paradigms," *Informatica*, vol. 29, no. 1, pp. 89-98, 2005.