International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 04 Issue: 12 | Dec-2017 www.irjet.net p-ISSN: 2395-0072

CERTAIN ANALYSIS ON TRAFFIC DATASET BASED ON DATA MINING ALGORITHMS

S. Revathi[1], Dr.A.Sumithra[2], S.Hebziba Jeba Rani[3], M.Vanitha[4]

[1,2,3,4] *Assistant Professor, Department of CSE, Sri Ramakrishna Institute of Technology, Tamilnadu, India*

Abstract - *Roadway traffic safety is a important consideration for transport agencies as well as normal peoples. Careful analysis on the roadway traffic data is important to give safe driving suggestions. In this paper we investigate the statistics analysis and data mining algorithms on the Traffic and Accident dataset as an attempt to address this problem. The attributes including collision manner, weather, surface condition, light condition, and drunk driver, Road condition, pavement maintenance were investigated. Association rules were discovered by Apriori algorithm, classification model was built by Naive Bayes classifier, and clusters were formed by simple K-means clustering algorithm. Certain safety driving suggestions were made based on statistics, association rules, classification model, and clusters obtained. Index Terms—Roadway fatal accidents, association, classification, clustering*

Keywords: ***Apriori, classification model, Naive Bayes classifier, association rules, clusters***

1.INTRODUCTION

Traffic monitoring at signals are very important nowadays because the number of vehicles increased and also there is a growth in traffic jams. The large number of vehicles is driving on the roadway every day and accidents could happen at any time anywhere. We all want to avoid accident and safe drive. To find out how to drive safer, data mining technique could be applied on the traffic accident dataset to predict some valuable information, thus give some driving suggestion.

The video cameras which are placed at signals are used for this purpose. There is a possibility for the video cameras to get spoiled by weather. Traffic security cameras would be damaged or ruined by heat, wind, rain, snow and ice. Current transportation environment can be improved in terms of traffic flow by integrating an intelligent computing methods for the roadside and probe vehicles. Intelligent Transportation Systems (ITS) probe vehicles with GPS tracker enable identification of traffic density and possible traffic jams. Updated traffic signal control which is connected to ITS server can reduce congestion. This paper gives a framework to mine GPS data for Intelligent Transportation Systems at Traffic Signals. Temporal analysis was also performed, using NLP techniques in order to detect the novelty of a tweet message

Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. It is considered one of the most important tool in information technology in the previous decades. Association rule mining algorithm and Apriori is a popular methodology to identify the significant relations between the data stored in large database and also plays a very important role in frequent itemset mining. A classical association rule mining method is the Apriori algorithm. Its main task is to find frequent itemsets, which is the method we use to analyze the roadway traffic data. Classification in data mining methodology aims at constructing a model (classifier) from a training data set that can be used to classify records of unknown class labels. The Naıve Bayes technique is one of the very basic probability-based methods for classification that is based on the Bayes' hypothesis with the presumption of independence between each pair of variable.

2.Related Works

Carlos Gutiérrez et al [1] Proposed to establish a computational framework, able to detect traffic related events in real-time, using social networks. The framework aims to be flexible enough to Twitter, but also to other social networks. As described in this paper, we investigated the real-time nature of Twitter, in particular for traffic event detection. Semantic analyses were applied to tweets to classify them into a positive and a negative class. They consider each Twitter user as a sensor, and set a problem to detect an event based on sensory observations. Name-entity recognition was used to extract location entities in a tweet message, and pinpoint those locations on a map. A temporal analysis was also performed, using NLP techniques in order to detect the novelty of a tweet message. From an implementation perspective, we developed a novel approach to notify people promptly of a traffic event. Despite the challenges associated with the real-time nature and length limitation that distinguish Twitter message from other social in this paper, real-time clustering approach which clusters messages from a stream of tweets. A tweet message tends to be more credible, if several users post similar messages in a very short period of time.

Liling Li et al [2] Suggested that association rule mining, and the classification, the environmental factors like roadway surface, weather, and light condition do not strongly affect the fatal rate, while the human factors like being drunk or not, and the collision type, have stronger affect on the fatal rate. From the clustering result we could

see that some states/regions have higher fatal rate, while some others lower. They may pay more attention when driving within those risky states/regions. Through the task performed, they realized that data seems never to be enough to make a strong decision. If more data, like non-fatal accident data, weather data, mileage data, and so on, are available, more test could be performed thus more suggestion could be made from the data.

Zhidan Liu et al [3] Presents a transport traffic estimation method which applies compressive sensing technique to achieve city-scale traffic estimation with only sparse traffic probes. The strong correlations among the road network is captured by an explicit model and further exploited to form a space basis that can sparsely represent the road traffic conditions. Through extensive trace-driven study and experiments, they validate the effectiveness of their traffic correlation model and show that our approach achieves accurate and scalable traffic estimation with only sparse probes.

Jie Xu et al [4] Proposed a framework for online traffic prediction, which discovers online the contextual specialization of predictors to create a strong hybrid predictor from several weak predictors. The proposed framework matches the real-time traffic situation to the most effective predictor constructed using historical data, thereby self-adapting to the dynamically changing traffic situations. They systematically proved both short-term and long-term performance guarantees for their algorithm, which provide not only the assurance that their algorithm will converge over time to the optimal hybrid predictor for each possible traffic situation but also provide a bound for the speed of convergence to the optimal predictor. Their experiments on real-world dataset verified the efficacy of the proposed scheme and showed that it significantly outperforms existing online learning approaches for traffic prediction.

Tianzhu Zhang et al [5] Stated a novel framework is proposed to mine semantic context information for intelligent video surveillance of traffic scenes. First, they introduce how to learn scene-specific context information from object-specific context information. Then, object classification is improved by combining of multiple features under a framework. Based on the learned information, they adopt it to improve object detection and tracking, and detect abnormal events. Experimental results validate that the semantic context information is effective to improve object detection, object classification, object tracking and abnormal event detection.

3. CONCLUSIONS

In the scope of this work, our analysis takes into account tweets posted by regional traffic agencies, where the problem concerned with the credibility of tweet messages is delimited and is not part of the presented work. Nevertheless, it is important to state that, traffic agencies act in the scope of this work as a provider and not as consumer of information. Into what this work is concerned, the main objective is not trying to solve all issues related with traffic condition of city rather than, to present an analysis which is capable to process the traffic dataset of the city and notify drivers about the status of the mobility network in real-time.

REFERENCES

[1] Carlos Gutiérrez, Paulo Figuerias, Pedro Oliveira, Ruben Costa, Ricardo Jardim-Goncalves. "Twitter Mining for Traffic Events Detection", Science and Information Conference (SAI), 2015, DOI: 10.1109/SAI.2015.7237170

[2] Liling Li, Sharad Shrestha, Gongzhu Hu. "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques", Software Engineering Research, Management and Applications (SERA), 2017 IEEE 15th International Conference, DOI: 10.1109/SERA.2017.7965753

[3] Zhidan Liu, Zhenjiang Li, Mo Li, Wei Xing, Dongming Lu. "Mining Road Network Correlation for Traffic Estimation via Compressive Sensing", Volume: 17, Issue: 7, July 2016,pp 1880 – 1893, DOI: 10.1109/TITS.2016.2514519

[4] Jie Xu, Dingxiong Deng, Ugur Demiryurek, Cyrus Shahabi, and Mihaela van der Schaar, "Mining the Situation: Spatiotemporal Traffic Prediction With Big Data Mining Semantic Context Information for Intelligent Video Surveillance of Traffic Scenes"

[5] Tianzhu Zhang, Si Liu, Changsheng Xu and Hanqing Lu. "Mining Semantic Context Information for Intelligent Video Surveillance of Traffic Scenes", IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, VOL. 9, NO. 1, FEBRUARY 2013