



CRIME ANALYSIS AND PREDICTION USING DATA MINING TECHNIQUES

****Rajkumar.S,*Sakkarai Pandi.M,*Soundarya Jagan.J,*Varnikasree.P**

*Students, Department of Computer Science ,
** Assistant Professor, Department Of Computer Science,
K.S.Rangasamy College Of Technology, Tiruchengode.

ABSTRACT

Crime analysis and prediction is a systematic approach for identifying and analyzing patterns and trends in crime. Our system can predict regions which have high probability for crime occurrence and can visualize crime prone areas. With the increasing advent of computerized systems. The main aim is that instead of focusing on cause of crime occurrences we are focusing mainly on crime factors. By using the concept of Data Mining, we can extract previously unknown useful information from an unstructured data. We can approach between computer science and criminal justice to develop a data mining procedure that can help solve crimes faster. Criminals also be predicted based on the crime data. Events of crime and illegal activities have increased in the past few years. We propose a system which can analyze, detect and predict various crime probability in a given region. To accomplish this, we obtain raw data from police department official website. On this pre processed data sets, by applying Naïve Bayesian algorithm we create a predictive model which analyze the data and helps to predict the trends of crimes for a given region in a future. With the aim of securing the society from crimes, there is a need for advance system and new approaches for improving the crime analytics for protecting their community. Accurate real time crime predictions help to reduce the crime rate. But remains challenging problem for the scientific community as a crime occurrences depend on many complex factors. The hidden relationship among the data which is further used to report and discover the crime patterns that is valuable for the crime analytics to analyze these crime networks by the means of various interactive visualization for crime prediction and hence is supportive in prevention of crimes. This probabilistic trend is also displayed in form of graphs for easy understanding of the police department. This paper explains various types of criminal analysis and crime prediction using several Data Mining techniques. Towards this goal, crime hotspot prediction has previously been suggested. Crime hotspot prediction leverages past data in order to identify crime hotspots, while ignoring the predictive power of other data such as urban

or social media data. Crime data analysts can help the law enforcement officers to speed up the process of solving crimes. Using the concept of data mining we can extract previously unknown, useful information from an unstructured data. Here we have an approach between computer science and criminal justice to develop a data mining procedure that can help solve crime faster.

Keywords

Data mining, machine learning, crime analysis, crime prediction.

1.INTRODUCTION

1.1Data Mining

Data Mining is the procedure which includes evaluating and examining large pre-existing databases in order to generate new information which may be essential to the organization. The extraction of new information is predicted using the existing datasets. Many approaches for analysis and prediction in data mining had been performed. But, many few efforts has made in the criminology field. Many few have taken efforts for comparing the information all these approaches produce. The police station and other similar criminal justice agencies hold many large databases of information which can be used to predict or analyze the criminal activity involvements in the society. The criminals can also be predicted based on the crime data. The paper presents the survey on the Crime analysis and crime prediction using several Data Mining techniques. The proposed Fraud detection methods in the telecom industry can be related to data mining techniques or machine learning algorithms as suggested by different research papers. Our main contribution in this research is to propose a new approach based on deep learning success in different classification tasks such object detection image recognition natural language processing and dimensionality reduction Deep learning algorithms use deep architectures of multiple layers to extract features from raw data through hierarchical progression of learned features using different layers from bottom navigating upwards without prior knowledge of ant rule, which it becomes more challenging when dealing with the huge amount of data, thus deep learning helps significantly to avoid feature engineering process which is time and resource consuming.

1.2.Crime Informatics

The crime analysis tasks can be a tedious process for the police or the investigation team to work with. The criminals when leaving the crime scene does leave some tracks which can be used as a clue to identify the criminals. The crime sequence and the patterns which several criminals follow when committing a crime make it easy for analyzing the crime. This process includes several procedures to be followed in order to identify the criminals and getting more information based only on the clues or information given by the local people. The criminal can be analyzed based on the information from the crime scene which is treated against the previous crime patterns and judging by the method which is implied to test and proceed with the information that can affect the prediction.

2.Existing System

These are enhanced query based layered approach for detecting SQL attacks and other web attacks, which is built from the ground up. Keeping in mind the various intricacies involved in such kind of attacks. The proposed system uses some knowledge base and query generation using the history of previous attacks and some java script feature which is determined by the user access level. Using the knowledge base detection of SQL attacks can be easily performed, it also maintains a list of some keywords, which make it easier to detect a large number of attacks at a faster pace the steps followed by in the proposed detection approach to identify the SQL injections above algorithm validates the input. That enhanced query based layered approach, for detecting SQL attacks and other web attacks. that enhanced query based layered approach, for detecting SQL attacks and other web attacks, which make it string using the initial knowledge base which stores all the frequent sql attacks of each category and is managed by the probabilistic approach if the new input string pattern already stored in the base, then it is declared as an sql attacks and a warning message will be generated automatically.

Drawbacks:

- Data quality can vary.
- Limits to question that can be asked and thus can be answered by further analysis.
- Comparisons between countries may not be valid because of the different ways data is collected and crime will be detected.
- Data may be missing in part for some years.

3. Proposed System

The proposed system is the technical challenges in designing and implementing the planner. First, the transformation operations are composable. The order of applying transformation operations also matters for performance and cost optimizations. The searching space for an optimal transformation sequence is huge. Second, the optimization is an online process and should be lightweight. This approach should find a good balance between the quality of the transformation sequence and the runtime overhead of the planner. Due to the huge space, a thorough exploration of the optimization space is impractical. Third, the planner should be able to handle different tradeoffs on the monetary cost and performance goals.

4. Module Description

Based on the pattern feature space generated by data processing we can carry out the last part of pattern recognition. The final output as this stage may be the type of object to which it belongs, or it may be the model number that is most similar to the object in the model database. A pattern, classification, or description, usually based on a collection of patterns that have been classified or described and this model is called a training set, and the resulting learning strategy is called supervised learning. Learning can also be unsupervised learning. In this sense, the system does not need to provide a prior knowledge of the pattern class, but it is based on the statistical laws of the patterns. In general, people use forward propagation networks for pattern recognition. Forward propagation is feedback that does not return to the input. Similar to what humans have learned from mistakes, neural networks can learn from their mistakes by feeding back information to the input.

4.1 Crime Frequent Hotspots

The first goal of our study was finding spatial and temporal criminal hotspots. We have successfully achieved this goal using Apriori algorithm on both Denver and Los Angeles datasets. We have extracted all the increasing patterns based on our predefined thresholds. We found that Denver has 62 interesting frequent patterns while Los Angeles has 59 patterns. The frequent item sets ordered by the location day of the week and the time period. With these different item sets. We are able to conclude the most likely crime location along with their frequent occurrence day and time.

4.2 Crime Prediction

The second target for our study was to predict the crime type that might occur in a specific location within particular time. The Bayesian classifier enabled us to reach this target with a reasonable accuracy. To predict an expected crime type, you need to provide four related features of the crime. The required features are: the occurrence month, the occurrence day of the week, the occurrence time and the crime location. All the features can be submitted in their nominal values.

4.3 Hotspots Demographics Analysis

After accomplishing our main goal by locating spatial and temporal criminal hotspots and predicting potential crime types, we applied some demographics analysis using Denver neighborhood demographics dataset. We wanted to further understand our model findings by studying the relationship between crime rate in each neighborhood and its description of demographics information. We found the peoples age and gender distribution vary between dangerous and safe locations. Specifically, dangerous neighborhood have more male while the safe neighborhoods have more female We have not found any relationship[between crime hotspots and peoples race distribution.

5. Conclusion

As a future extension of our work , We plan to apply more classification models to increase crime prediction accuracy and to enhance the overall performance. It is also a helpful extensions for over study to consider the income information for neighborhoods in order to see if there iare relationships between neighborhoods income level and their crime findings. Further more, we want to study other crimes datasets from ne cities along with their demographics datasets.

Acknowledgements

We acknowledge DST-File No.368.DST-FIST (SR/FIST/College-235/2014 dated 21 – 11 – 2014) for financial support and DBT-STAR-CollegeScheme-ref.no:BT/HRD/11/09/2018 for providing infrastructure support.

6.References

- [1] Akoglu, Leman, and Christos Faloutsos.(2015) "Anomaly, event, and fraud detection in large network datasets." Proceedings of the sixth ACM international
- [2] Bengio, Yoshua, Nicolas Boulanger- Lewandowski, and Razvan Pascanu.(2016) "Advances in optimizing recurrent networks." Acoustics, Speech and Signal Processing (ICASSP), IEEE International journal on. IEEE.
- [3] Bergstra, James, and Yoshua Bengio.(2015) "Random search for optimization." Journal of Machine Learning Research 13-281-305.
- [4] Boureau, Y-Lan, Jean Ponce, and Yann LeCun. "A theoretical analysis of feature pooling in visual recognition." Proceedings of the 27th international journal on machine learning.
- [5] Chang, Yi-Chun (2017) "Mining the Networks of Telecommunication Fraud Groups using Social Network Analysis." [6]Ehab Hamdy, Ammar Adl, Aboul Ella Hassanien, Osman Hegazy and Tai-Hoon Kim,(2015) "Criminal Act Detection and Identification Model", Proceedings of 7th International journal Advanced Communication and Networking, pp. 79-83.
- [7] Glorot, Xavier, and Yoshua Bengio. "Understanding the difficulty of training feedforward neural networks." Aistats. Vol. 9.
- [8] He, Kaiming,(2016) "Deep residual learning for image recognition" Proceedings of the IEEE journal on computer vision and pattern recognition.
- [9]Hilas, Constantinos S., Paris A. Mastorocostas, and Ioannis T Rekanos.(2016) "Clustering of Telecommunications UseProfiles for Fraud Detection and Security Enhancement in Large Corporate Networks: A case Study." Applied Mathematics & Information Sciences 9.4.
- [10] Jyoti Agarwal, Renuka Nagpal and Rajni Sehgal,(2015)"Crime Analysis using Means Clustering", International Journal of Computer Applications, Vol. 83, No. 4, pp. 1-4.

