

A Real Time Crime Intelligence System for Suspect Prediction

Thejaswini N (M.Tech)

*Department of Computer Science and Engineering
Vidyavardhaka College of Engineering
Mysuru, Karnataka, India.*

Shraddha C, Assistant Professor

*Department of Computer Science and Engineering
Vidyavardhaka College of Engineering
Mysuru, Karnataka, India.*

Abstract- In recent years the data mining is data analyzing techniques that used to analyze crime data previously stored from various sources to find patterns, suspects and trends in crimes. In additional, it can be applied to increase efficiency in solving the crimes faster and can be applied to automatically notify the crimes. However, there are many data mining techniques. In order to increase efficiency of crime detection, it is necessary to select the data mining techniques suitably. This paper reviews the literatures on various data mining applications, especially applications that applied to solve the crimes. Survey also throws light on research gaps and challenges of crime data mining. In additional to that, this paper provides insight about the data mining for finding the patterns, suspects and trends in crime to be used appropriately and to be a help for beginners in the research of crime data mining.

Keywords- Crime Intelligence; C4.5 Algorithm; Classification Rules; Data Mining; Decision Trees.

I. INTRODUCTION

Crime prevention and detection become an important trend in crime and a very challenging to solve crimes. Several studies have discovered various techniques to solve the crimes that used to many applications. Such studies can help speed up the process of solving crime and help the computerized systems detect the criminals automatically. In addition, the rapidly advancing technologies can help address such issues. However, the crime patterns are always changing and growing. The crime data previously stored from various sources have a tendency to increase steadily. As a consequence, the management and analysis with huge data are very difficult and complex. To solve the problems previously mentioned, data mining techniques employ many learning algorithms to extract hidden knowledge from huge volume of data. Data mining is data analyzing techniques to find patterns and trends in crimes. It can help solve the crimes more speedily and also can help alert the criminal detection automatically.

Data Mining, also popularly known as Knowledge Discovery in Databases (KDD), refers to the nontrivial extraction of implicit, previously unknown and potentially useful

information from data in databases. While data mining and knowledge discovery in databases (or KDD) are frequently treated as synonyms, data mining is actually part of the knowledge discovery process.

Proposed system is applicable in the field of crime. Proposed system includes modeling of crimes for finding suitable algorithms to detect the crime, precise detection, data preparation and transformation, and processing time. Proposed system identifies crime behavior, crime predicting, precise detection, and managing large volumes of data obtained from various sources. Proposed system is an automation for complaints registration, crime suspect prediction based on the previous crime details collected from various sources.

2. DATA MINING

The removal of several hidden patterns from a large database which appears useful is called data mining. It is a strong technology with substantial potential to help and knowing the world and explain natural phenomenon. Over the years, people have been gathering and analyzing data organizations concentrate on the most valuable information in their database. However, gradually new technologies have begun to play a vital role to handle the storage, analysis and processing of data. Specially, the advent of computer technology has revolutionized the way in which data are managed. This new method of searching through the data as well as the keen interest to learn from data has brought disciplines like that of data mining. Some researchers also noted that data mining is valuable to discover relevant and useful information from huge data stored in the database through building computer programs that sort through the database automatically, seeking meaningful patterns [18]. The opportunity for the application of data mining has increased tremendously as databases grew extremely and new machine with searching capabilities evolved. Looking into this data has introduced various theories, observations, and approaches that will help in understanding the law and knowing the natural world [2].

A. The Data Mining Process

Data mining is not all about the use of software [2]. It is a process that involves series of steps to transform data prior to mining, evaluate and interpret modelling results. It is the

process of discovering relevant patterns in a large amount of data that can describe or tell us about past events in a way that the modelled results can be employed to predict the future. According to [2], the most frequently used data mining steps are; identifying the source of the data, preparing data for analysis, building and training a computer model and evaluating the computer model.

B. Classification Rule Mining and Decision Trees

Classification is a commonly used data mining technique; it uses a large population of records for classification to create a model from a set of pre-classified examples. This technique makes use of decision tree or neural network-based classification algorithms. The most common task in data mining is to build models to predict the class of an object based on its attributes. Classification trees can have binary or unary branches. Most times the tree structure in classification trees have binary branches, when we split the data in two ways it will result in a better separation. The classification process of data involves learning and classification. In learning, the classification algorithm analyzes the data to be trained. In classification, the accuracy of the classification rules estimates the test data. The classifier training algorithm makes use of pre-classified examples to determine the set of parameters needed for proper discrimination. Classification is also called supervised learning; supervised learning is a process where network can learn from target vector containing the desired output from a pair of input vector. The prediction next time would be closer to the correct answer if the learning algorithm can take the differences between the correct output and the prediction of neural networks.

C. How Decision Trees Work

According to [2], decision tree can be used to classify and predict the dividing records in the database to small sets in relation to values of other fields. The decision tree construction starts at the root node where you take actions from the root node you split the remaining nodes constantly irrespective of the decision tree learning algorithm. The output is a decision tree with each branch having a decision made at each step. Classification task is a rule of the decision being made. It expresses if-then' explicitly compared to neural networks due to its strength and popularity. Decision tree works with computer-generated rules which can provide an explanation for its action unlike neural networks.

II. METHODOLOGY

Proposed system is an automation for complaints registration, crime suspect prediction based on the previous crime details collected from various sources.

A. Data Collection and Description

The data used in this research was collected from an extensive dataset of crimes in Chicago ,by city of Chicago. The data collected consists of crime records for about ten years, that is, from 2001-2017,

The data used in this work consists of records. which were described by the attributes: programme, sex, offence, weapon used, expulsion period and level. A fraudulent record forms a transaction. Theoretically, we cannot have a fraudulent record. The behavior of a person appears to be the only way crime can be investigated and this can be done by checking out the relating attributes, which was built on forensic data then developed on record attributes.

TABLE 1: Description of the feature selection

FEATURE	DESCRIPTION
Name	Names of people and organisations.
Location	Location of occurrence of crime.
Type of crime	Crime types such as Murder, Kidnapping, Robbery etc.
Date	Date of occurrence of crime.
Time	Time of occurrence of crime.
Period of day	Period of the day as to Morning, Afternoon or Evening.
Day of week	Day of the week when the crime is reported to have occurred.

B. Information retrieval and Feature extraction

The investigation officer keys in a brief description of the crime, containing details about the location, date of occurrence of crime and keywords related to the type of crime as an input. The given input is classified based on the type of crime by applying feature selection technique CRF++. CRF++ [6] is a simple, customizable, and open source implementation of Conditional Random Fields (CRFs) for segmenting/labeling (classifying) sequential data. In this study, CRF++ is trained and applied to identify the features such as Location, Day, Date, Time and the type of crime as to Murder, Kidnap or Robbery.

C. Retrieval of relevant cases

Based on the features extracted from the given input the relevant case reports are drawn from a database that contains different reports about the different type of crime cases. A consolidated summary of the relevant cases are displayed based on C4.5 Algorithm.

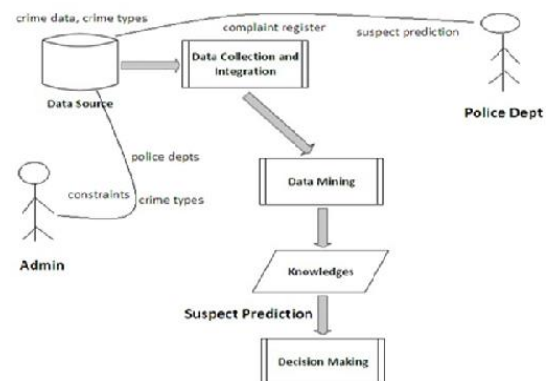


Fig.1: Modules used to build the system

D. C4.5 Algorithm

C4.5 algorithm, used to create Univariate Decision Trees and also Multivariate Decision Trees, their process to classify instances using more than one attribute per node in the tree.

The algorithms used to create Univariate DT's and multivariate DT's. This one, called C4.5, is based on the ID3 algorithm, that tries to find small (or simple) DT's. We start presenting some premisses on which this algorithm is based on the inference of the weights and tests in the nodes of the trees.

Construction:-

Step 1: Scan the dataset (storage servers)

Step 2: for each attribute a, calculate the gain [number of occurrences]

Step 3: Let a_best be the attribute of highest gain [highest count]

Step 4: Create a decision node based on a_best– retrieval of nodes where the attribute values matches with a_best.

Step 5: recur on the sub-lists [list of patient] and calculate the count of outcomes[Stages] – termed as subnodes. Based on the highest count we classify the new node.

III. EXPERIMENTS AND RESULT

Proposed system identifies crime behavior, crime predicting, precise detection, and managing large volumes of data obtained from various sources. Proposed system is automation for complaints registration, crime pattern prediction based on the previous crime details collected from various sources. The suspect is predicted based on the number of times he/she attempted it by calculating feature count based on constraints.

Sl.No	Constraint Name	Value
1.	WeaponUsed	Knife
2.	Reason	Asset

Fig: Results: Suspect prediction

IV. CONCLUSION

This project reviews the data mining applications, especially applications that applied to solve the crimes. In addition to that, this project provides insight about the data mining for finding the crime suspects. To solve the problems, data mining techniques employ many learning algorithms to extract hidden knowledge from huge volume of data.

V. FUTURE WORK

As in this paper when the dataset is very large it is computationally takes large amount of time to find and calculate the feature count which can be reduced in the future and it is difficult to identify the new people who involved in crime which will be enhanced in future.

VI. REFERENCES

- [1]. Sathyadevan, M. Devan, and S. Surya Gangadharan, "Crime analysis and prediction using data mining," in Networks Soft Computing (ICNSC), 2014 First International Conference on, Aug 2014, pp. 406–412.
- [2]. T. Pang-Ning, S. Michael, and K. Vipin, "Introduction to Data Mining", 1st ed. Pearson, 2005.
- [3]. S. Kaza, Y. Wang, and H. Chen, "Suspect vehicle identification for border safety with modified mutual information," in Proceedings of the 4th IEEE International Conference on Intelligence and Security Informatics, ser. ISI'06.
- [4]. V. Vaithyanathan, K. Rajeswari, R. Phalnikar, and S. Tonge, "Improved apriori algorithm based on selection criterion," in 2012 IEEE International Conference on, Dec 2012, pp. 1–4.
- [5]. C. Chu-xiang, S. Jian-jing, C. Bing, S. Chang-xing, and W. Yun-cheng, "An improvement apriori arithmetic based on rough set theory in Circuits, Communications and System" (PACCS), 2011 Third Pacific-Asia Conference on, July 2011, pp. 1–3.
- [6]. W. Huang, M. Krneta, L. Lin, and J. Wu, "Association bundle – a new pattern for association analysis," in Data Mining , Sixth IEEE International Conference on, Dec 2006, pp. 601–605.
- [7]. A. Thammano and P. Kesisung, "Enhancing k-means algorithm for solving classification problems," in Mechatronics and Automation (ICMA), IEEE International Conference on, Aug 2013, pp. 1652–1656.
- [8]. Y. Zhao and G. Karypis, "Evaluation of hierarchical clustering algorithms for document datasets," in Proceedings of the Eleventh International Conference on Information.
- [9]. R. Marikhu, J. Moonrinta, M. Ekpanyapong, M. Dailey, and S. Siddhichai, "Police eyes: Real world automated detection of traffic violations," in Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2013 10th International Conference on, May 2013, pp. 1–6.
- [10]. W. Cheng, X. Ji, C. Han, and J. Xi, "The mining method of the road traffic illegal data based on rough sets and association rules," in Intelligent Computation Technology and Automation (ICICTA), 2010 International Conference on, vol. 3, May 2010, pp. 856–859.