

Crime rate prediction using k-means

Rehnuma Reza Deepty(Roll:11)

Albina Alam(Roll:15)

Tanzim Kabir(Roll: 29)

26 February 2019

System Description

In present scenario criminals are becoming technologically sophisticated in committing crime and one challenge faced by intelligence and law enforcement agencies is difficulty in analyzing large volume of data involved in crime and terrorist activities therefore agencies need to know technique to catch criminal and remain ahead in the eternal race between the criminals and the law enforcement. So appropriate field need to chosen to perform crime analysis and as data mining refers to extracting or mining knowledge from large amounts of data, data mining is used here on high volume crime dataset and knowledge gained from data mining approaches is useful and support police forces.

To perform crime analysis appropriate data mining approach need to be chosen and as clustering is an approach of data mining which groups a set of objects in such a way that object in the same group are more similar than those in other groups and involved various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. In this paper k means clustering technique of data mining used to extract useful information from the high volume crime dataset and to interpret the data which assist police in identify and analyze crime patterns to reduce further occurrences of similar incidence and provide information to reduce the crime. In this paper k mean clustering is implemented using open source data mining tool which are analytical tools used for analyzing data. Among the available open source data mining suite such as R, Tanagra, WEKA, KNIME, ORANGE, Rapid miner. k-means clustering is done with the help of rapid miner tool which is an open source statistical and data mining package written in Java with flexible data mining support options. Also for crime analysis dataset used is Crime dataset, an offences recorded by the police in Bangladesh by offence and police force area from 2017-2018.

Crime analysis is defined as analytical processes which provides relevant information relative to crime patterns and trend correlations to assist personnel

in planning the deployment of resources for the prevention and suppression of criminal activities. It is important to analyze crime due to following reasons:

1. Analyze crime to inform law enforcers about general and specific crime trends in timely manner
2. Analyze crime to take advantage of the plenty of information existing in justice system and public domain

Crime rates are rapidly changing and improved analysis finds hidden patterns of crime, if any, without any explicit prior knowledge of these patterns. The main objectives of crime analysis include:

1. Extraction of crime patterns by analysis of available crime and criminal data
2. Prediction of crime based on spatial distribution of existing data and anticipation of crime rate using different data mining techniques
3. Detection of crime

Data mining in the study and analysis of criminology can be categorized into main areas, crime control and crime suppression. De Bruin et. al. introduced a framework for crime trends using a new distance measure for comparing all individuals based on their profiles and then clustering them accordingly. Manish Gupta et. al. highlights the existing systems used by Indian police as e-governance initiatives and also proposes an interactive query based interface as crime analysis tool to assist police in their activities. He proposed interface which is used to extract useful information from the vast crime database maintained by National Crime Record Bureau (NCRB) and find crime hot spots using crime data mining techniques such as clustering etc. The effectiveness of the proposed interface has been illustrated on Indian crime records.

Nazlena Mohamad Ali et al. discuss on a development of Visual Interactive Malaysia Crime News Retrieval System (i-JEN) and describe the approach, user studies and planned, the system architecture and future plan. Their main objectives were to construct crime-based event; investigate the use of crime based event in improving the classification and clustering; develop an interactive crime news retrieval system; visualize crime news in an effective and interactive way; integrate them into a usable and robust system and evaluate the usability and system performance and the study will contribute to the better understanding of the crime data consumption in the Malaysian context as well as the developed system with the visualization features to address crime data and the eventual goal of combating the crimes.

Sutapat Thiprungsri examines the application of cluster analysis in the accounting domain, particularly discrepancy detection in audit. The purpose of his study is to examine the use of clustering technology to automate fraud filtering during an audit. He used cluster analysis to help auditors focus their efforts when evaluating group life insurance claims.

A. Malathi et al. look at the use of missing value and clustering algorithm for a data mining approach to help predict the crimes patterns and fast up the process of solving crime. Malathi. A et. al. used a clustering/classify based model to anticipate crime trends. The data mining techniques are used to analyze the city crime data from Police Department. The results of this data mining could potentially be used to lessen and even prevent crime for the forth coming years. Dr. S. Santhosh Baboo and Malathi. A research work focused on developing a crime analysis tool for Indian scenario using different data mining techniques that can help law enforcement department to efficiently handle crime investigation. The proposed tool enables agencies to easily and economically clean, characterize and analyze crime data to identify actionable patterns and trends.

Kadhim B. Swadi Al-Janabi presents a proposed framework for the crime and criminal data analysis and detection using Decision tree Algorithms for data classification and Simple K Means algorithm for data clustering. The paper tends to help specialists in discovering patterns and trends, making forecasts, finding relationships and possible explanations, mapping criminal networks and identifying possible suspects. Aravindan Mahendiran et al. apply myriad of tools on crime data sets to mine for information that is hidden from human perception. With the help of state of the art visualization techniques we present the patterns discovered through our algorithms in a neat and intuitive way that enables law enforcement departments to channelize their resources accordingly.

Sutapat Thiprungsri examine the possibility of using clustering technology for auditing. Automating fraud filtering can be of great value to continuous audits. The objective of their study is to examine the use of cluster analysis as an alternative and innovative anomaly detection technique in the wire transfer system. K. Zakir Hussain et al. tried try to capture years of human experience into computer models via data mining and by designing a simulation model.