May-June 2020
ISSN: 0193-4120 Page No. 3658-3660

Crime Detection Using Machine Learning Method

M. Raja Suguna[1], A. Kalaivani[2]
[1]Assistant Professor, [2]Associate Professor,
[1,2]Department of Computer Science, Saveetha School of Engineering,
Saveetha Institution of Medical and Technical Sciences, Chennai
[1]suguna.raj89@gmail.com, [2]kalaivania.sse@saveetha.com

Article Info
Volume 83
Page Number: 3658-3660
Publication Issue:
May-June 2020

Article History
Article Received: *19 August 2019*
Revised: *27 November 2019*
Accepted: *29 January 2020*
Publication: *12 May 2020*

Abstract

Crime is the action that constitutes the serious offence against a person or a state and punishable by law. Crime Pattern Analysis assist the police personnel and detective to predict the next relative occurrence of the crime. This is done by analyzing the massive recorded crime data over a geographical region for a particular period of time. K-Means Clustering Algorithm and Geocoding is used to extract Spatial features for different genre of crime recorded over a particular Geographical region. The Spatial feature extracted is used for training the Supervised Learning Algorithm such as Support Vector Machine and Logistic Regression to predict the future crime region and time.

Keywords: *Crime, Crime Pattern Analysis k-Means Clustering, Geospatial Feature.*

1. Introduction

Crimes are the threat to the mankind. In recent days, there is a large volume of crime data available. Advancement of Artificial Inelligence Techniques urges the researchers to analyze the large volume of crime data to predict the future crime and to identify the crime prone areas which supports police Department to the larger extend. The work cannot predict who is going to be offender and the victim, instead it reveals the relative next crime region and the time of occurrence of the crime. These Predictions makes the Police Department to extend their patrol in the identified hotspots (predicted future crime region) at the predicted time to enforce law and order.

The different Categories of crime are (i) Violent Crime, which includes murder and rape (ii) robbery (iii) traffic Violence (iv) alcohol-related (v) property crime. The Model created is concentrated on above mentioned genre of crimes. First, K-Means Clustering Algorithm is used to find the potential Clusters. These Cluster stores the location of the crimes along with the time. These regions are called hotspots, which are identified as the cluster with a greater number of records of crime. These features are then used for predict the future crime by Support Vector Machines and Logistic Regression.

2. Related Works

R. Iqbal et al[1] used WEKA, an open source Data Mining tool to compare two Classification Algorithm Decision Tree and Naïve Bayes Algorithm for predicting the crime for the US Crime dataset from LEMAS Survey.T. Beshah et al [2] compared three data mining three Classification algorithm K-Nearest Neighbor, Naïve Bayes and Decision stump for Ethiopia road accident data and got their prediction accuracy ranging between 79% to 81%.Most Prediction models in crime prediction used the spatial features from the crime data to find the hotspots, where the crime rates are more than the average levels. Q. Zhang et al [3] used Linear Discriminative Analysis and statistical methods as a spatio-temporal model for predicting the hotspots along with the KNN Algorithm. F. K. Bappee et al [4] used density based Spatial Clustering algorithm to extract the spatial feature from crime data and combine with the geocoding to predict the hotspots by making use of Open Street Map. Chainey, et al [5] and Nakaya, T et al [6] used Kernel Density Estimation to predict the hotspots using spatio-temporal features of crime data. KeivanKianmehr [7] used Support Vector Machines for predicting thr crime hotspots. Chung-Hsien Yu et al [8] compares the other Machine Learning Models with Support Vector Machine for Crime hotspot prediction.

In our Proposed work ,spatial Features are extracted by using K- Means Clustering Algorithm and geocoding .Classification algorithm such as Support Vector Machines, Logistic Regression used to predict the Future crime region and time.

TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 3658-3660

3. Data Analysis

a) Data Source

The original Sanfranciso Crime Dataset is downloaded from the opensource kaggle platform. The dataset contains more than fifteen lakh crime records with 13 columns for the crime recorded during the year 2016.The attributes under which the crime data recorded incident number, crime category, Description about the crime, day of week, Date, time, District, Resolution -kind of punishment given to the crime, Address, x-latitude of the crime location, y – longitude of the crime location, location – exact location name, pd ID.

The columns, Description and Resolution is removed from the datasetas the work is on prediction of the future category and the time of crime. The dataset is filtered for having the crime record of the category (i) robbery (ii) traffic Violence (iii) alcohol-related (iv) property crime.

b) Engineering Statistical Features

Geocoding Technique is used to represent the exact location of the crime by combining descriptive details from the crime data such as latitude, longitude, address of the crime location. This output the spatial representation of the location of the crime. Geocoding uses GIS to give exact location of crime. ArcGIS API of Python is used in this work which uses spatial reference of the geocoder (4326).The locations are visualized in WebMap and WebScene Component Provided by API.

The location output of the geocoding also be grouped by Category such as hotels, amusement park, hospital, bus stop etc., Hence the Spatial representation of the crime region is done by both using exact location of GIS and also by using the category of location in Map.

c) Finding Hotspots

Creation of hotspot is the important step in crime Analysis. Hotspot is the data pertaining to the time and the location of the geographical region. K-Means Clustering algorithm is used to identify the hotspots of the crime data. The algorithm clusters the geographic region having the higher occurrence of the crime of same category. After identifying the hotspots, the centroids of each hotspot are examined to find the next relative centroid of the future crime. Hotspots are identified for the four different classes of crime category mentioned in the previous section.

d) Prediction of Crime Using Machine Learning Algorithm

Classification algorithm used are Support Vector Machines and Logistic Regression. N-fold cross Validation Method is carried on the preprocessed Sanfranciso Crime Data for training each Classification Model. Accuracy is calculated for the Two Models with the raw features (actual attributes of Data) and also with the engineered attributes such as Geocoding and the hotspots along with raw features of the data separately.

Table 1 shows the Model Accuracy of SVM and Logistic regression for the four different category of crime.

Table1: Result for Classification Accuracy

Crime Category	Support Vector Machines		Logistic Regression	
	Raw Features	Spatial Features	Raw Features	Spatial Features
Robbery	59.63	67.8	56.89	62.54
Traffic Violence	63.26	69.22	60.21	63.2
Alcohol-related	88.5	90.1	83.25	86.52
Property Crime	76.32	83.21	74.26	79.32

4. Conclusion

In this work, two spatial features such as Geocoding and Hotspot are explored for the four different categories of crime in crime data. These two features relatively increase the prediction accuracy of Classification Algorithm such as Support Vector Machines and Logistic regression. The future direction of work is to predict for all category of crime and with the increasing accuracy of predicting hotspots of crime.

References

[1] R. Iqbal, M. A. A. Murad, A. Mustapha, P. H. ShariatPanahy, and N.Khanahmadliravi, "An experimental study of classification algorithms forcrime prediction," Indian J. of Sci. and Technol., vol. 6, no. 3, pp. 4219-4225, Mar. 2013

[2] T. Beshah and S. Hill, "Mining road traffic accident data to improvesafety: role of road-related factors on accident severity in Ethiopia," Proc.of Artificial Intell. for Develop. (AID 2010), pp. 14-19, 2010.

[3] Q. Zhang, P. Yuan, Q. Zhou, and Z. Yang, "Mixed spatial-temporalcharacteristics based crime hot spots prediction," IEEE 20th Intl. Conf. onComput. Supported Cooperative Work in Des. (CSCWD), Nanchang,China, May 2016.

[4] F. K. Bappee, A. S. Júnior, and S. Matwin, "Predicting crime using spatialfeatures," Can. AI 2018: Advances in Artificial Intel.-Lecture Notes inComput. Sci., vol. 10832, pp. 367-373, Springer, Mar. 2018.

[5] Chainey, S., Tompson, L., Uhlig, S.: The utility of hotspot mapping for predictingspatial patterns of crime. Security Journal 21(1), 4{28 (2008)

[6] Nakaya, T., Yano, K.: Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation andscan statistics. T. GIS 14(3), 223{239 (2010)

May-June 2020
ISSN: 0193-4120 Page No. 3658-3660

[7] KeivanKianmehr and Reda Alhajj. "Effectiveness ofsupport vector machine for crime hot-spots prediction".In: Applied Artificial Intelligence 22.5 (2008), pp. 433–458.

[8] Chung-Hsien Yu et al. "Crime forecasting usingdata mining techniques". In: Data Mining Workshops(ICDMW), 2011 IEEE 11th International Conferenceon. IEEE. 2011, pp. 779–786.