International Journal of Pure and Applied Mathematics
Volume 118 No. 22 2018, 1777-1781
ISSN: 1314-3395 (on-line version)
url: http://acadpubl.eu/hub
Special Issue

ijpam.eu

SENTIMENT ANALYSIS AND GEOGRAPHICAL ANALYSIS FOR ENHANCING SECURITY

Namratha M[1], R Hariharan[2], Nimmi K[3], J SangeethaPriya[4]
[1,2,3,4]Assistant Professor, [1]Department of CSE, BMS College of Engineering, Bangalore,
[2]Department of Information Technology,
Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology,
[3]Department of CSE, SCMS School of Engineering and Technology,
[4]Department of IT, Saranathan college of Engineering ,
[1]namratha.july@gmail.com, [2]hharanbtech@gmail.com, [3]menimmicse@scmsgroup.org
[4]priyamail.kumar@gmail.com

Abstract: Enhancing security around our general public has been the major challenging issue .This paper is a point by point study about the endeavours that have been made to enhance the forecast of the crime intensity utilizing machine learning and big data to help the citizens to remain in a sheltered zone. In this manner, the sort of study discussed here will help reveal the crime rate of an area progressively. This paper focuses on both the setimental analysis as well as geogrpahic analysis of an area to give more accurate results regarding the type of the crime in a particular area.

Keywords: Machine learning, Sentiment Analysis, Tweets, KDE(Kernel Density Estimation)

1. Introduction

Crime is a forbidden act that is punishable by authorities .it can be any social nuisance which cause defined as an act harmful not only to the individual and also to the community ,society etc.

The issues of the crime are growing in certain cities of the world. So, there is a need for some technology which can predict the occurrences of crime which means automated crime detection is much needed in the society. This will help the police departments in great extent.

The study and analysis of crime is done using computer technology, legal strategies along with tactics targeting what type of crime ,for example murder, robbery etc, leads to the prediction of digital crime, digital terrorism, an information warfare.

The technology used here is geographical analysis[1] along with sentimental analysis of online social networking sites which help in the detection of crime patterns. Twitter is one of the online social networking and micro-blogging feature that enables users to post short description referred to as "tweets". These updates may convey important information. Here a system is designed which extract tweets from cities that are regarded to be dangerous cities in the country.

A geographic analysis and the sentiment analysis[2][3] of data uncovered a connection between tweets and crimes that occurred in the corresponding cities. Sentiment analysis techniques were implemented on these tweets to analyze the intensities of crime in particular location. These processes need to be identified and their performance can then be improved using data from previous executions.

2. Literature Survey

Nowadays with the help of statistical learning techniques, computers can learn and identify the patterns in the given data. The most common applications uses of machine learning are Data Mining, Natural Language Processing (NLP), Image and Speech Recognition, Medical diagnosis, finance, transportation, retail and social media services industry. It is being set as the pillar of future civilization.[4][5][6]

Twitter as of now serves roughly 140 million worldwide users posting a consolidation 340 million messages (or tweets) per day. Bollen Stated that, "a tweet is a microscopic, temporally-authentic instantiation of sentiment". Since tweets are brief, people in general opinion can be effectively investigated. Twitter likewise provides the feature of Re-tweet (RT), which allows users to give comments of comments. Large scale analysis of tweets might provide insights that are not apparent within a single field. Thus we need to channel in lights on our needs.

Sentiment Analysis is the process of 'computationally' determining whether a piece of writing is positive, negative or neutral. It's also known as 'opinion mining', finding the attitude or opinion of the

user. Sentiment analysis[7] also is used to monitor and analyse social phenomena, for the spotting of possibly risky circumstances. The quick development of online networking has built the interest in sentiment. Various forms of online expressions (e.g., opinions-like reviews, ratings, and recommendations) have turned out to be a significant source of information to the organisation to deal with their reputations. The challenge of detecting crime patterns lies in geographically analyzing the hot spot areas[8][9] in the city and then performing sentiment analysis on the tweets of the city to find red-alerts areas in real-time. The geographical analysis is done using the static data source and sentimental analysis is done with real-time tweets. This study helps use in predicting the crime using both real-time and static data sources more precisely and accurately.

In this study we do the following:

1. Collect the data for geographical analysis from various open sources. Use the data collected and analyse it using the KDE[10][11] (Kernel Density Estimation) which gives crime intensity distribution and helps us determine hot spot areas. Based on crime intensities we plot graph, explained in the Related Works in the next session.

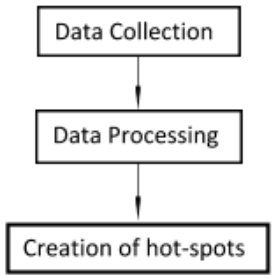

Figure 1. flow diagram of system

2. Collect the real-time data from twitter API. Analyze textual content in Twitter data by using sentiment analysis to obtain the polarity of tweets [12] and their trends in different neighbourhoods. Tweets are collected and filtered based on hot-spot areas (output of KDE). Analysis is carried out to get the sentiment of the tweet. We categorize them into positive, negative, very positive, very negative and neutral. Besides sentiment polarity, we were also interested in the trend of sentiment within each neighbourhood. More detailed description of sentimental analysis is explained in related work in next session.

3. For the prediction, we first calculate the estimated crime density of each neighbourhood on the certain period by using the KDE method mentioned above. Then we process tweets on the that period to calculate the polarity score and trend index. The logistic regression model uses estimated density and twitter features to predict the likelihood of incidents occurring on that period in its neighbourhood. The graph plotted thus shows the intensity of each crime in the city and hence helps notifying the citizens about the appropriate measures to be taken.[13][14]

3. Related Work

Here, techniques to be used are included and are following:

i) Kernel Density Estimation (KDE):An approach to crime prediction is hotspot mapping, based on the assumption that crime predictions that the locations of past events are good predictors of future events. This approach generally includes two categories: methods that rely on aggregate crime data and analysis of discrete crime event locations. [15][16]The end result is visual representation of crime points, thereby creating continuous risk surface. Hot spot identification depends on the size of the geographical area of concern, location and addresses to streets, blocks and neighbourhoods. The computation equation for KDE is as follows:

$$x(1,t)(p) = k(p,h) = \frac{1}{Ph}\sum_{j=1}^{P} K\left(\frac{\vee\, p - p(j)\,\vee}{h}\right)$$

We calculate crime density at each point p. We set the following parameters:

'h' as bandwidth parameter

'P' as total number of crime incidents that occurred

'j' indicates single crime point from the data of crime period

'K' is standard normal density function

$\|.\|$ is Euclidean norm

'p(j)' actual crime point.

Methods used for identifying crime hot spots are divided into two general categories: Aggregate incident location and analysis of point-patterns.

ii) Sentiment Analysis: In this examination, we propose to study the feature-based [17] opinion summarisation of criminal tweets. The assignment is performed in two stages: First, recognize the highlights of the tweets that people have tweeted. Second, for each element we distinguish what number of tweets have positive or negative opinions. The particular surveys that express these conclusions are appended to the element. This facilitates browsing of the surveys by potential clients. In terminology identification, there are essentially two techniques for finding terms in corpora: symbolic approaches that depend on syntactic depiction of terms, to be specific noun phrases, and factual methodologies that make us of the way that the words making a term tend to be discovered near each other at reoccurring. The system performs the summarisation in two main steps:

'Feature Extraction' and 'Opinion orientation identification'. Feature Extraction, given the data sources, the framework initially downloads all the tweets, and places them in the survey database. The feature extraction work, first concentrates "hot" highlights that many individuals have communicated their opinions in their tweets, and after that finds the infrequent once. Opinion orientation identification, function takes the generated features and compresses the feelings of the component into two classifications, positive and negative. Frequent features generation, this step is to find features that people are most interested in. In order to do this, we use association rule mining to find all frequent itemsets. An association rule[18] is an implication of the form X --> Y, where X-->I, Y-->I, and X --> Y = Ø. The rule X-->Y holds in D with confidence c. If c% of transactions in D that support X also support Y. The rule has support s in D if s% of the transactions in D contain X?Y. Mining frequent occurring phrases, each bit of data extracted above is put away in dataset called a transaction set/document. We at that point run the association rule miner, which depends on the Apriori algorithm. This algorithm works in two stages, initial step is to discover all frequent item sets from an arrangement of exchanges that satisfy a user-specified minimum support. Feature Purning Compactness pruning, method checks features that contain at least two words, which we call feature phrases, and remove those that are likely to be meaningless. Redundancy Purning, we focus on removing redundant features that contain single words. Opinion Sentence orientation Identification, after opinion features have been identified, we determine the semantic orientation (i.e., positive or negative) of each opinion sentence. This consists of two steps: (1) for each opinion word in the opinion word list, we identify its semantic orientation using a bootstrapping technique and the WordNet and (2) we then decide the opinion orientation of each sentence based on the dominant orientation of the opinion words in the sentence.

iii) Prediction using sentimental polarity and historical crime record as variables: Our goal is to predict future crime on certain areas of the city. To predict the future crime, we are interested in the trend of sentiment within each neighbourhood along with sentiment polarity. Intuitively, consecutive periods of positive or negative sentiments might cause a greater risk of crime. So, to measure the trend, we create a trend index. The motivation of the algorithm comes from the pattern of stock costs. The data available from twitter depends on time. This means that each observation of a dataset has a time tag attached to it. This kind of data is known as 'time-series data'. The current ways to deal with this issue, thrown in two gatherings: Feature Filters and Feature wrappers.[19] The former are independent of the modelling tool that will be used after the feature selection phase. They basically try to use some statistical properties of the features (ex. Correlation) to select the final set of features. The latter approach include the modelling tool in the selection process. They complete an iterative search process where each iteration is a candidate set of features is tried with the modelling tool and the respective results are recorded. The decision of models was predominantly guided by the way that these techniques are well known by their ability to handle highly nonlinear regression problems. Detailed approaches to this domain would necessarily require a large comparison of more alternatives. Two such alternatives are listed be: Artificial Neural Network and Support-Vector machine. (1) Artificial Neural Networks(ANNs) are frequently used in financial forecasting because of their ability to deal with highly nonlinear problems. ANNs are formed by a set of computing units (the neurons) linked to each other. (2) Support Vector Machine, modelling tools that, as ANNs, can be applied to both regression and classification tasks. SVMs have been seeing expanded consideration from various research groups in perspective of their fruitful application to a few areas and furthermore their solid hypothetical foundation. The fundamental idea behind SVMs is that of mapping the original data into another, high-dimensional space, where it is conceivable to apply straight models to get an separating hyper plane, for instance, isolating the classes of the problem, in the case of classification tasks.

4. Conclusion

Public Security is major challenging issue. The machine learning techniques can be efficiently integrated in the field of crime investigation and thus help us to take precautionary measure before the incidents occur. It will be helpful to prevent major crime at very fast manner. Machine learning techniques easily find and analyze how crime is going to happen.

References

[1] HARDI. M. PATEL, RIPAL PATEL "Enhance algoritham to predict a crime using datamining", Journal of Emerging Technologies and Innovative Research (JETIR) ,Volume 4, Issue 04, April 2017 Pgno:257-259

[2] Rui Xia , Jie Jiang, and Huihui He "Distantly Supervised Lifelong Learning for Large-Scale SocialMedia Sentiment Analysis" IEEE transactions on affective computing, vol. 8, no. 4, october-december 2017 ,480-491

[3] Harpreet Kaur, Veenu Mangat ,Nidhi ,"A survey of sentiment analysis techniques", International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) feb 2017, 921-925

[4] Metin Bilgin, Izzat Fatih ,"Analysis on Twitter data with semi-supervised Doc2Vec, Computer Science and Engineering (UBMK), 2017 International Conference onYear: 2017,pgno:661-666

[5] Ankit Kumar Soni ,"Multi-lingual sentiment analysis of Twitter data by using classification algorithms", Computer and Communication Technologies (ICECCT), feb :2017,

[6] Ms.A.M.Abirami, Ms.V.Gayathri ,"A survey on sentiment analysis methods and approach" 2016 IEEE Eighth International Conference on Advanced Computing (ICoAC) Jan 2017,pg.no:72-76

[7] Poornima Nedunchezhian, Shomona Gracia Jacob "Social Influence Algorithms and Emotion Classification for Prediction of Human Behavior: A Survey , " Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM), feb :2017,pg.no: 55-60

[8] WalaaMedhat[a]AhmedHassan[b]HodaKorashy[b] "Sentiment analysis algorithms and applications: A survey", Volume 5, Issue 4, December 2014, Pages 1093-1113

[9] Mohammad A. Tayebi, Martin Ester, Uwe Glasser , Patricia L. Brantingham "CRIMETRACER: Activity space based crime location prediction," IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Aug 2014,pg.no:472-480

[10] Anthony J. Corso, Gondy Leroy, Tucson Abdulkareem Alsusdais, , Tucson Abdulkareem Alsusdais "Toward Predictive Crime Analysis via Social Media, Big Data, and GIS", iconference 2015

[11] S.V.Manikanthan and T.Padmapriya "Recent Trends In M2m Communications In 4g Networks And Evolution Towards 5g", International Journal of Pure and Applied Mathematics, ISSN NO: 1314-3395, Vol-115, Issue -8, Sep 2017.

[12] S.V. Manikanthan, T. Padmapriya "An enhanced distributed evolved node-b architecture in 5G tele-communications network" International Journal of Engineering & Technology (UAE), Vol 7 Issues No (2.8) (2018) 248-254.March2018.

[13] S.V. Manikanthan, T. Padmapriya, Relay Based Architecture For Energy Perceptive For Mobile Adhoc Networks, Advances and Applications in Mathematical Sciences, Volume 17, Issue 1, November 2017, Pages 165-179