November-December 2019
ISSN: 0193-4120 Page No. 4830 - 4837

Small Area Crime Prediction using Deep Neural Net

Lydia J Gnanasigamani, Seetha Hari*
Vellore Institute of Technology, Vellore
VIT-AP, Amaravathi

Article Info
Volume 81
Page Number: 4830 - 4837
Publication Issue:
November-December 2019

Article History
Article Received: *5 March 2019*
Revised: *18 May 2019*
Accepted: *24 September 2019*
Publication: *23 December 2019*

Abstract:
In this work, we have leveraged the power of Deep Neural Networks to make day to day crime count prediction in a small area of the city boundaries. We use Chicago city data to make the predictions. The crime data is augmented with weather, transportation and census data. We also study the effect of these augmented data on the accuracy of the model. We split the crime counts into 5 bins based on their counts. The model then predicts the most probable crime bin for each of the small areas on a day to day basis. The model is also trained on the variations in temporal and spatial characteristics of the crime prediction. The outcomes of the experiments demonstrate that the Deep Neural Networks are very effective in predicting crime count of a small geographical area.

Keywords: *Crime, Deep Learning, Deep Neural Networks, Framework, Natural Language Processing, Prediction*

1. INTRODUCTION

From ancient times humans have been fascinated by the unknown. And the future is the greatest unknown. We do not know what might happen in the future. There have many prophets, prophecies and seers who have tried to predict some of the events that may happen after their time. Likewise, crime has been always part of human life. Wherever humans settled as a society, crime was always there. Today we deal with many different types of crime and with limited resources we need to know where to deploy them to prevent the majority of crimes. Crime data historically has been in the hands of the police officers and they have been extrapolating trends manually. But the increase in computing power, the reducing costs of hardware and open data policies have made the data available to every day. So there have been many types of research which leverage the computing methods to find future trends.

Many data mining techniques have been applied to find crime patterns and prediction. Statistical methods like regression were used to predict the crime rates in future years. But this was done purely based on historical data. Then machine learning methods like supervised and unsupervised learning was used for crime prediction. But the prediction time scales were at a year or longer period. Multivariate regression improved the prediction power. The addition of other data which are known influencers of crime like weather, socio-economic indicators, and the transient population was not considered as these add to the existing features and the computing requirement rose very high. The increase in computing powers in recent days have given way to deep learning, a comparatively young

November-December 2019
ISSN: 0193-4120 Page No. 4830 - 4837

field in artificial intelligence. Deep learning has achieved great success in multiple domains such as computer vision, speech and classification problems. Its application to prediction problems has not been studied to a great extent.

In this paper, we propose to use deep learning neural networks to predict crime in a smaller geographical area. This prediction is done on a day to day basis. The remaining sections of the paper are organized as follows,

First, we discuss the existing literature on crime forecasting, then we introduce deep neural networks. Finally, we go through the training and testing of the network.

2. LITERATURE REVIEW

Crime prediction methods have changed over time with the growth in technology and the availability of data. Prediction methods are a way of summarizing past experiences into future expectations [1]. Unpredictability is the nature of the crime. But by analyzing the previous crimes and history of criminality we can try to say the probability of occurrence of crime. Most of the studies in crime have been oriented towards environmental criminology to determine the factors that may encourage crime. Using statistical methods it has been determined that the offenders mostly involve in criminal activities near their home [2]. The West Midlands police data was used to check the relation between offender age, their address and the crime location. It was shown that as the offenders get older the crime areas tend to get nearer to their home. Most of the activities of the offenders were within a mile of their home [3].

Crime prediction has also been attempted with various machine learning and data mining algorithms. With the widespread use of mobile phones, it has been used to study human behaviours [4]. This behavioural data, and the reported crime data along with other demographic data like the weather were used for borough level crime predictions in the city of London [5]. Among all the ensemble algorithms used, the random forest algorithm gave the best performance. Tham and Hofer studied the possibility of predicting if an individual may involve in criminal activity or not and to provide intervention and prevent crime in such cases [6]. The study data were derived from the Sweden criminal data sets and other surveys that were conducted to identify the level of living of the people in Sweden. They found a strong correlation between early childhood risk factors and crime. A person exposed to risk factors in their early childhood had a major possibility of being arrested due to a criminal offence between the ages 18 to 36. But even then we cannot say who may get involved in multiple crimes. Thus they concluded even with socio-economic data it is not quite possible to predict when and if an individual may get involved in crime. Researchers have also represented crime data as a graph and have tried to find patterns and predict future crime rates using them [7]. A data mining tool using k-means and k-Nearest Neighbor (kNN) clustering methods was developed by Tayal et al [8]. The tool uses data from seven major cities in India and uses it to predict the crime rate. Prediction of future crime trends using DBSCAN method has been proposed [9]. DBSCAN and K-means both were used to analyze crime data from Indian crime database and both the algorithms had their pros and cons. A new algorithm a cross between k-means and DBSCAN has been proposed and its performance has been evaluated. The new algorithm shows better accuracy than the individual algorithms. Almanie et al. have used the Apriori algorithm to find patterns from crime hotspots using crime data from Denver and Los Angeles [10]. They have also used the crime trend using Naïve Bayes and decision tree. Naïve Bayes had an accuracy of about 52%

TEST Engineering & Management
November-December 2019
ISSN: 0193-4120 Page No. 4830 - 4837

and decision tree methods accuracy was 42%. Both of these were not satisfactory enough for prediction.

A predictive model was constructed to predict hit and run cases using tweets [11]. The tweets about hit and run cases from a news feed are automatically gathered, then information from them is extracted using natural language processing. Then a generalized linear regression model was applied to predict the future hit and run cases. The prediction was successful to an extent. As most of the twitter data have GPS tagging, the weather for a particular location is found using weather API [12]. Once the tweets about crime are collected, then sentiment analysis is performed on them. Then the prediction is done on the resultant data using Kernel Density Estimation (KDE) method. Twitter data has also been used as augmented data for crime data to enhance the prediction abilities of the KDE model [13]. The model showed a markable increase in crimes like stalking and criminal damage. Crime data along with urban indicators such as elderly population, child labour, female population, GDP, illiteracy, income, employment are used to predict crime [14]. Random forest generator is used and the results show that unemployment and illiteracy are the most influential indicators for homicide.

Hotspots are places where the incidents of crime are much more than other places. Prediction of hotspots for the Northeast US police department was done using crime data and spatial information [15]. The whole area was split into uniform grids and the hotpots in each grid were identified for burglary using various classifiers. SVM and neural nets gave the best prediction of hotspots. Today's advances in hardware and computation powers have made neural networks more accurate and powerful. In our paper, we use a deep neural network to predict crime.

3. NEURAL NETWORK MODELS

We use the feed-forward neural network in this work and use it get a prediction for each day. The preparation of data for the model is detailed in the next section, but the basic format is explained here. The cities are split into grids using the police beats for Chicago. This information is already available in the UCR data portal for the two states. Each grid contains an additional set of characteristics for each day. One record in the data contains all the attributes for a day for each grid. The prediction is done on a day to day basis in line with the data. The prediction count for each grid is done by using intervals like 0-10, 11-20 etc. We follow the supervised learning methodology for our model. That is the prediction interval for each grid is known. The models need to be trained before they can be used for prediction. The mathematical concepts related to training and testing a feed-forward model is given in the next subsection. But the training can be described briefly using the following steps:

1. Randomly initialize the network weights
2. For each instance in the data set
   a. Read each instance and input it to the neural network
   b. Do the forward pass with the current set of weights to get the prediction
   c. Find the difference between the actual and the predicted output using a loss function
   d. Perform backpropagation to disseminate the errors to all the individual nodes.
   e. Recalculate the weights using the loss function to find the gradient
3. Find the average of the gradient
4. Recalculate the weights using the average gradient and a learning rate
5. Repeat the steps from 2 to 4 until the weights are stabilized.

November-December 2019
ISSN: 0193-4120 Page No. 4830 - 4837

After sufficient epochs/iterations, the model stabilizes. The random weights are recalculated to form accurate measures that create precise predictions. In Deep Neural Networks there is no requirement for the feature selection process. The network determines the important features for the given task on its own accord. This is the reason that deep neural nets are able to handle a very large amount of attributes than the previous machine learning methods.

2.1. Deep Neural Network

Artificial neural networks (ANN) are processing systems motivated by the natural neural systems in humans and animals. These systems "learn" using the inputs and do not require a framework of rules. ANN consists of input layers, output layers and the intermediate layers known as the hidden layers as shown in Figure 1. The input layer is the first layer of the network and the number of nodes in the input layer is in parallel with the number of attributes in the dataset. The output layer has the number of nodes equal to the possible number of predictions. The number of hidden layers is resolved by trails. The input layer also has a bias which generally takes the value 1 enabling the use of network even when there is no input. Increasing the number of hidden layers (depth) amplifies the networks ability to learn. But it may also cause the network to overfit, meaning unique cases may be considered as normal and the network may be tuned to that.

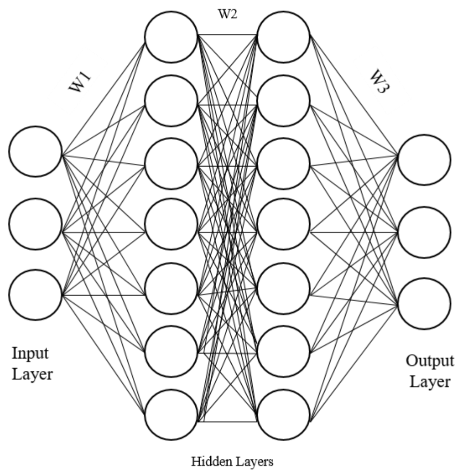

Figure 1 Deep Neural Network

Every node is connected to every other node in the next layer by weight. Each instance in the dataset is fed to the input layer and the values are fed forward to the other layers until it reaches the output layer. Every node takes the weighted sum of its inputs calculated by performing matrix multiplication and the result is passed to a non-linear activation function. This then becomes the input of another node in the next layer of the network. For the hidden layers, we have used ReLU (Rectified Linear Unit) activation function and for the output layer, we use the softmax activation function. Softmax is used because it provides the probability for each different output or for each grid in our case.

The behaviour of hidden layers can be mathematically given as

$$a_{t,h+1} = f(b + a.w) = f\left(b + \sum_{i=1}^{n} a_i w_i\right)$$

where$a_{t,h}$is the input to the layer $h+1$, f is the non-linear function that represents the network and is expressed in terms of the weight vector w and the bias vector b. ReLU activation function used in the hidden layers is defined as $f = \max(0, v)$where $r = wa + b$.

If M is the number of grids or beats, then for a network with H hidden layers, the softmax

November-December 2019
ISSN: 0193-4120 Page No. 4830 - 4837

activation is given by $w_H a_{t,H} + b_H$ and we get a vector of probabilities p for each grid G. where the probability of bin $b \in \{1 \dots G\}$ is

$$p_{th}^{b}(a_t) = \frac{\exp\left(v_{th}^{b}\right)}{\sum_{b=1}^{G} \exp\left(v_{th}^{b}\right)}$$

where v_{th} is the output vector and a_t is the input. The loss for a softmax layer is

$$S_{th} = \sum_{b=1}^{G} l_{th}^{b} \log p_{th}^{b}(a_t)$$

where $l_{th}^{b} = 1$ if the count of the crime in the instance a_t falls in the grid b and 0 otherwise. This loss is calculated for each beat and for every input instance. Then we take the derivates of this loss function and the loss is backpropagated through the network using the chain rule. The gradients of the lower layer are calculated and the weights in that layer are updated.

4. DATA PREPARATION AND TRAINING

We use the crime dataset from the data portal for Chicago, https://data.cityofchicago.org. In Chicago, there are 274 beats. We also can download the shapefile with the beats from the portal. Using the beats information in the dataset, we can map the crimes to a particular beat. There are many different fields in the dataset such as block information, the location where the crime occurred such as residence or street, the type of crim, the UCR code of the crime and binary data like if an arrest was made for that crime or not. There are lots of different crime types. We separate them into 12 crime types that cover all the different types of crime. Like theft, vehicle theft, other thefts are all group under theft. For each day each beat the crimes are summed up. The summed up numbers are binned according to their count according to their bins. This number forms the input to our neural network. Figure 2 shows the distribution of counts for a day for 4 beats. There is not much difference in the crime count for each hour of the day. We consider the crime data available from the year 2001 and only non-residential data. Figure 3 shows the distribution of assault, burglary, narcotics and homicide type crimes in Chicago.

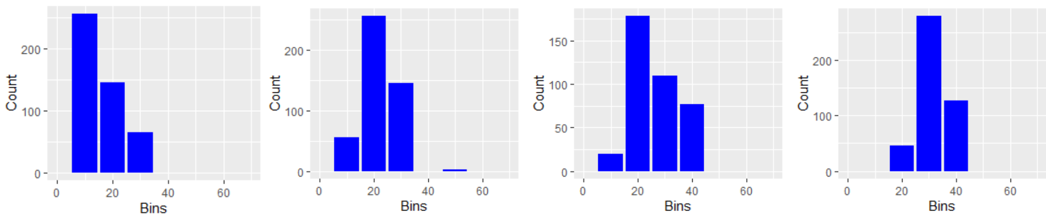

Figure 2 Counts for a day for four different beats in Chicago

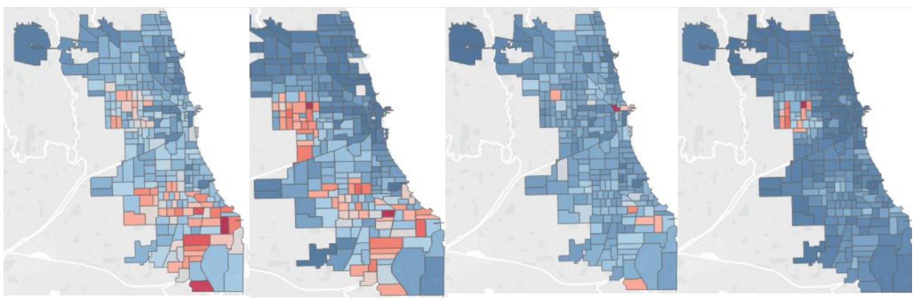

Figure 3 Distribution of assault, burglary, homicide and narcotics in Chicago

November-December 2019
ISSN: 0193-4120 Page No. 4830 - 4837

We also use transportation dataset from the same portal. The transportation dataset gives details about public transport such as buses and trains with their source, destination and intermediate stops. These stops are also associated with the latitude and longitude information, which we use to map it to the beats. They also provide information on the number of commuters per train and bus route. The number of entries at each railway station is also available. Those these stations and numbers are not specific to the beats, they do touch each of the beats thereby affecting them. The stations are mapped to the grids or beats and the number of commuters is totalled up for each beat.

The crime dataset provides accurate information of the time, date along with the spatial information, the latitude and longitude, for each reported crime. We use the spatial data, date and time information to get the weather information for the particular beat from https://www.ncdc.noaa.gov. We use this site to find the nearest weather station and map it to a beat. Not all the weather stations data are complete. So we use the minimum and maximum temperature for the day, humidity, snowfall if any, depth of snow.

The census data from the 2010 census is used to get the socioeconomic features such as income, education, age, percentage of each race etc. Similar to transportation information this information are mapped to the beats and they also form part of the input layer.

The feature breakdown is as follows. For each beat, we have binned the crime types into 12 bins. We use 6 weather features such as minimum temperature for the day, maximum temperature for the day, average temperature, precipitation, snowfall, and depth of snowfall. Total commuter numbers for each beat calculated from the transportation data set. The transportation dataset also provides day information such as special holidays, weekends, weekdays etc. For the crime, the aggregated count of crimes in each beat for the 12 bins of the crime type is calculated.

For training neural networks, the data is split into training and test data randomly. 90% of data is used for training and the rest 10% for testing. But for a time-series data like crime where we day by day information in a progressive order, it doesn't make sense to take data randomly. We cannot use the later date data for training and test using earlier crime data. So the data is split into batches using the year. In a given year, the first 90% of data is used for training and 10% is used for testing. This is repeated for all the years. The weights from the initial model are passed on to the next model (for the next year) and so on. This process is shown in figure 4.

5. EXPERIMENTS

We have designed three experiments to test our models. In the first experiment, we use all the datasets and attributes that we have discussed in the previous section. For the deep ANN, we used three hidden layers with ReLU activation. This was then followed by layers equal to the number of beats. All these layers are softmax layers, as in they use softmax activation function which returns the prediction corresponding to the beat. Table 1 shows the accuracy results of all the experiments.

Table 1. Prediction Results

	Experiment 1	Experiment 2	Experiment 3
Chicago	72.3	68.6	71.1

November-December 2019
ISSN: 0193-4120 Page No. 4830 - 4837

In the second experiment, we try to see the impact of the additional data that we have used in addition to the crime dataset. We methodically remove the features of the transportation, weather and census data and see the effect these datasets have in the accuracy of prediction. The accuracy drops for the removal of every dataset. Removing the census data has the most impact on accuracy. Census data is the most steady data of all the additional datasets. And the dropin the accuracy of the model on the removal of these data from the input proves that crime and socio-economic attributes have a very strong correlation. The decrease in, accuracy, is the least for the weather dataset. This may be because of the limited geographical space that we have considered. As it is only a city and not a big geographical area like a state or country, it fairly possible for it to have a homogenous weather condition across the city. Transportation data also contributes significantly to the accuracy. It provides the count of the number of persons that pass through the beat. And when there are more people going through any place the higher the probability of crime.

In the next experiment, we studied the clauses that affect the model accuracy. There are some beats which always have low value for the crime, between 0 to 10 counts of occurrence per day. For these beats, the accuracy prediction was always better than for other beats. Likewise, the prediction was better for the beats which had consistently higher crime count more than 25 crime counts per day. For these beats was also the accuracy was fairly good. It was next to accuracy than the low crime beats. The performance of beats that had the beat counts in between the low and high values had high variability. Next, we check for the impact of weather attributes on the accuracy of the model. The temperature does not have any effect on accuracy. Again it may be because our study area is city and mostly has the same temperature across the city. Then we consider the snow precipitation and the accuracy of the model. On the days with a normal precipitation level of 70% to 75% there is no change in the accuracy. But the accuracy of the model reduces on the days with very high snow precipitation levels. We think this may be due to very few data available for these very high precipitation levels. Transporation data contains the station details and the number of commuters that use the station. It also indicates if the day is a normal weekday, a weekend or a holiday. Now the models show better accuracy for prediction during weekdays than weekends. We think this may be because the steady commuter traffic during the weekdays and weekends bring unusual traffic. The accuracy further decreases during holidays as there are more activities then and that brings a lot of unexpected and unusual traffic which leads to unpredictable circumstances.

6. CONCLUSION

We have created a Deep Neural Network model that predicts the next day crime count for a small geographical area. We have also used weather, transport and census data and have studied the impact of all these factors on the model's prediction accuracy. The model fares well for most of the cases. We can still try to vary the hidden layers and check for the improvement in the accuracy. As we have used the police beats as a boundary, the prediction from this model can be used to identify the beats with predicted higher crime counts and deploy police resources accordingly. This may act as a preventive measure and the crime counts will go down in the future. Also, the incorporation of GIS and crime maps will increase the accuracy of the prediction. With crime maps, we can identify crime concentration points within a beat. We can use Google APIs to identify nearby places like bars,

TEST Engineering & Management

November-December 2019
ISSN: 0193-4120 Page No. 4830 - 4837

transport hubs, shopping centres which may be considered a crime attracter. Once these places are identified we can start using preventive actions like increasing CCTV coverage, deploying more uniformed policemen, patrol cars. We also could try predicting the changes in crime counts for smaller regions which is more challenging as we need more localized and accurate data.

REFERENCES

[1] D. M. Gottfredson, "Assessment and Prediction Methods in Crime and Delinquency," in *Contemporary Masters in Criminology*, Springer US, 1995, pp. 337–372.

[2] V. Grover, R. Adderley, and M. Bramer, "Review of Current Crime Prediction Techniques," in *Applications and Innovations in Intelligent Systems XIV*, Springer London, 2007, pp. 233–237.

[3] D. K. Rossmo, "Geographic Profiling," Dec. 1999.

[4] Y. A. De Montjoye, J. Quoidbach, F. Robic, and A. Pentland, "Predicting personality using novel mobile phone-based metrics," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013, vol. 7812 LNCS, pp. 48–55.

[5] A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi, and A. Pentland, "Once upon a crime: Towards crime prediction from demographics and mobile data," in *ICMI 2014 - Proceedings of the 2014 International Conference on Multimodal Interaction*, 2014, pp. 427–434.

[6] H. Tham and H. von Hofer, "Individual Prediction and Crime Trends," *Eur. J. Criminol.*, vol. 6, no. 4, pp. 313–335, Jul. 2009.

[7] S. Sathyadevan, M. S. Devan, and S. Surya Gangadharan, "Crime analysis and prediction using data mining," in *1st International Conference on Networks and Soft Computing, ICNSC 2014 - Proceedings*, 2014, pp. 406–412.

[8] D. K. Tayal, A. Jain, S. Arora, S. Agarwal, T. Gupta, and N. Tyagi, "Crime detection and criminal identification in India using data mining techniques," *AI Soc.*, vol. 30, no. 1, pp. 117–127, 2014.

[9] M. A. Assistant Professor, S. Santhosh Baboo Reader, and P. Graduate, "An Enhanced Algorithm to Predict a Future Crime using Data Mining," 2011.

[10] T. Almanie, R. Mirza, and E. Lor, "Crime Prediction Based on Crime Types and Using Spatial and Temporal Criminal Hotspots," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 4, pp. 01–19, Jul. 2015.

[11] X. Wang, M. S. Gerber, and D. E. Brown, "Automatic crime prediction using events extracted from twitter posts," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2012, vol. 7227 LNCS, pp. 231–238.

[12] X. Chen, Y. Cho, and S. Y. Jang, "Crime prediction using Twitter sentiment and weather," in *2015 Systems and Information Engineering Design Symposium, SIEDS 2015*, 2015, pp. 63–68.

[13] M. S. Gerber, "Predicting crime using Twitter and kernel density estimation," *Decis. Support Syst.*, vol. 61, no. 1, pp. 115–125, 2014.

[14] L. G. A. Alves, H. V. Ribeiro, and F. A. Rodrigues, "Crime prediction through urban metrics and statistical learning," *Phys. A Stat. Mech. its Appl.*, vol. 505, pp. 435–443, Sep. 2018.

[15] C. H. Yu, M. W. Ward, M. Morabito, and W. Ding, "Crime forecasting using data mining techniques," in *Proceedings - IEEE International Conference on Data Mining, ICDM*, 2011, pp. 779–786.