International Journal of Research in Engineering, Science and Management
Volume-3, Issue-11, November-2020
https://www.ijresm.com | ISSN (Online): 2581-5792

Crime Branch Associated with Big Data and Data Mining

Brinda Jagadeesh[1*], Divya S. Naik[2], Aishwarya Gopinath[3], D. Pradeep Kumar[4]
[1,2,3]*Student, Department of Computer Science and Engineering, M. S. Ramaiah Institute of Technology, Bengaluru, India*
[4]*Assistant Professor, Department of Computer Science and Engineering, M. S. Ramaiah Institute of Technology, Bengaluru, India*
*Corresponding author: brindaja@gmail.com

***Abstract*: Data analytics is the art of analyzing raw massive data in order to draw conclusions or to forecast or to keep a track of current trends exists in a particular field.**

***Keywords*: Big data analytics [BDA], Crime pattern, Crime trend prediction, Data analytics [DA], Data mining, Predictive model.**

1. Introduction

Big data is a platform to deal with areas like analyzing, extraction of information from complex data sets, in other words can be regarded as predictive analytics, user behavior analytics. Data mining is a tool that helps in tracing the information from massive data sets [1].

In these present days, both the population and number of crimes occurring are directly proportional. So, our paper mainly emphasizes on the extraction or tracing of crime pattern based on previous and present data that are available in all police station departments.

Now-a-days with the advancement of technology the human beings are able to get all the information needed to his/her work at one place, hence traditional method of searching a crime case by using file system and then tracing the details from it consumes a lot of time, and may not be able to solve the entire case. So, making an attempt to trace the crime pattern from previous data by using big data analytics and data mining is the main work frame of this paper. The crimes can be traced by taking various cities of the country and implementing the appropriate model to get the number of crimes (here the prophet model is considered).

2. Related Work

A. BDA

Big data analytics and data mining superposed with predictive models can help us extract the insights from heterogeneous and massive data sets. Big data analytics not only helps a computer science graduate but rather it can help all private and public sectors and government sectors can also use these methods for their benefits [2].

B. Crime mining, trend prediction

Crime detection or Crime mining is the method of solving by merging crime branch along with data mining.

A class of crimes refers to a particular area(location) i.e., the number of crimes that are said to be committed would be more in that region, those are predicted using spatial plots by a map at the police station department [3].

Crime trend prediction: Crime trend prediction is the ability to predict future crimes, and thereby preventing the social crimes in the worst affected areas.

C. Working method

1) Pre-processing the data

Attributes to consider,

a) Crime No.: No of crimes that is likely to occur or has already occurred.
b) Date and Time: Interval of time and date (if previously available)
c) Crime Patt: Type of crime committed (robbery/harassment)
d) Theory of crime: Summary of crime recorded
e) Day type: Week or per month
f) Police dept. num: Nearby police station number
g) Approach to resolve: Action plan
h) Place of crime: Location like street description
i) Domestic: If crime is committed by a family
j) Action: Arrested or not.

2) Attributes to consider for plots

a) X axis: Longitudes of location of crime
b) Y axis: Latitudes of location of crime

3. Proposed Method

A. Prophet model

Prophet is a non-proprietary infographic forecasting library. Its goal is to give high quality outcome for decision making without having the necessity for the user to have knowledge about infographic forecasting and how it works. The prophet model breaks down infographic forecasting into holidays, trend

and seasonality in the equation.

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \quad (1)$$

For the function trend g(t),

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma) \quad (2)$$

$$a_j(t) = \begin{cases} 1, & if\ t \geq s_j, \\ 0, & otherwise. \end{cases} \quad (3)$$

where k is the growth rate, δ is the rate adjustment, m is the offset parameter, and γ is set to $-s_j\delta_j$ to make the function continuous.

For seasonality,

$$s(t) = \sum_{n=1}^{N} (a_n \cos(\frac{2\pi nt}{P}) + b_n \sin(\frac{2\pi nt}{P})) \quad (4)$$

For holiday,

$$Z(t) = [l(t \in D_1, \ldots, l(t \in D_L))] \quad (5)$$

$$h(t) = Z(t)k \quad (6)$$

where D_i is the set of past and future dates for holidays, $k \sim$ Normal$(0,v^2)$, l is an indicator function representing whether time t is during holiday i, Z(t) is the regressor matrix.

B. Neural Network Model

In this neurons are associated to each other, which in network are called nodes. The inputs are collaborated using

$$z_j = b_j + \sum_{i=1}^{n} \omega_{i,j} x_i \quad (7)$$

The hidden layer will modify the input above using a nonlinear function by

$$s(z) = \frac{1}{1 + e^{-z}} \quad (8)$$

C. LSTM model

They are advantageous in predicting because they have the capacity of accommodating the state along with identifying patterns across time series shown below,

$$f(t) = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (9)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (10)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (11)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (12)$$

where, f_t is a sigmoid function to indicate whether to keep the previous state, C_{t-1} is the old cell state, C_t is the updated cell state, W_f, W_i, and W_C are the previous value in each layer, h_{t-1} and x_t,is the input value, b_f, b_i,and b_c are constant values, i_t decides which value will be used to update the state, C_t stands for the new candidate values.

4. Experimental Results

Time series can showcase a range of patterns, and it is advantageous when it is disintegrated into numerous. For evaluation of performance in crimes, RMSE and spearman correlation are used

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \tilde{Y}_i)^2} \quad (13)$$

$$\rho_{X,Y} = \frac{cov(X, Y)}{\sigma_X \sigma_Y} \quad (14)$$

where Y_i and Y are the true values; $\tilde{Y}_i$ and X are predicted values; cov is the covariance; σ_X is the standard deviation of X, and σ_Y is the standard deviation of Y.

5. Conclusion

We conclude that Prophet and LSTM models provide us with a moderate result. This can also predict the optimal attributes for Prophet and LSTM models.

References

[1] Hassani, Hossein, Xu Huang, Emmanuel S. Silva, and Mansi Ghodsi, "A review of data mining applications in crime." Statistical Analysis and Data Mining: The ASA Data Science Journal 9, no. 3 (2016): 139-154.

[2] J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility," IEEE Trans. Electron Devices, vol. ED-11, no. 1, pp. 34–39, Jan. 1959.

[3] Archenaa, J., and EA Mary Anita. "A survey of big data analytics in healthcare and government." Procedia Computer Science 50 (2015): 408-413.

[4] Nath, Shyam Varan. "Crime data mining." In Advances and Innovations in Systems, Computing Sciences and Software Engineering, pp. 405-409. Springer, Dordrecht, 2007.