A Hybrid GCN and LSTM Structure Based on Attention Mechanism for Crime Prediction

Jinming Hu

School of Information Network Security, People's Public Security University of China, Beijing 102628, China

*Corresponding Author.

Abstract

Globalization has been the major contributor to economy boom. While at the same time, it has stimulated the development of crime method as frequent cross-border communication allowed. With the improvement in big data and prediction system of policing work, it has become a new research field to establish an efficient crime prediction model, by which police departments could clamp down on criminal activities more accurately. Besides, this model will be quite beneficial for commanding and dispatching police force thus to improve work efficiency. This paper proposes a combination model, which uses Long Short-Term Memory Network (LSTM) and Graph Convolutional Network (GCN) to predict crime rate and takes advantage of the Attention mechanism to improve the experimental result. By extracting the spatio-temporal characteristics of crimes and increasing the proportion of typical feature, it can not only predict crime quantity, but also detect the degree of crime risk in each region. A rolling forecast of crime data for about three years in the Boston of the United States shows that our model has good prediction performance.

Keyword : *Crime prediction, graph convolutional network, long short-term memory network, spatial-temporal, attention*

I. Introduction

Nowadays, with the improvement of people's living standard and technological progress in transportation, urbanization is much more rapid than ever before.[1]. However, as people enjoying a colorful and abundant life, the ways of committing crimes are also diversifying. Around the world, the cooperation between police and IT firms are becoming frequent to fight against crime under the strong support of technology. The basic jobs of the police include preventing, stopping, and cracking down on criminal activities, maintaining public order and so on [2]. Among them, crime prevention is the priority. Therefore, improving both the technological method and accuracy for crime prediction has become the goal of technology enterprise who join hands with the police. With the continuous development of big data and the popularity of deep learning, crime prediction has undoubtedly become a research frontier [3]. There are different types of crime around the world, namely larceny, drugs, robberies, assault, and telecommunication fraud. The police have been fighting against emerging telecom fraud in recent years. It is also an indication of continuous development in crime technology. Therefore, the establishment of crime prediction system should be highlighted.

Larceny generally refers to the act of illegal possession of public property or other people's property. Although theft rate has been gradually reduced over years, it is never a problem we could take lightly. Due to its low cost, convenient and quick way of committing crimes, and the trend of gang crime, cracking down on larceny cases still have a long way to go [4]. Previous study on crime data analysis demonstrated that holiday, season, and climate change are positively correlated with larceny rate. That is to say, larceny is predictable and the establishment of warning system is of great significance.

Drug crime can be broadly separated into 2 categories, namely, drug production and drug trafficking. There is a worldwide-consensus on combating drug crime. Although marijuana is legal in some states of the United States, drug legalization is just an exception [5]. As far as drug prediction is concerned, the characteristics are traceable as location drug manufacturing tends to be on border areas, and drug trafficking happened in in most major entertainment venues. It can be concluded that the spatial characteristics is a significant factor in drug cases forecast.

Assault cases generally include common assault and aggravated assault cases. Such cases are hard to predict since they erupted all sudden [6]. The current method to predict such cases is depending on previous data system and analyze the quality of public security in specific region. There is a direct relation between assault and public order, so predicting such cases are of great significance public security.

In order to extract the temporal and spatial characteristics of various cases and estimate the number of crimes in time and space automatically, a new model based on Attention mechanism is proposed in this paper. Firstly, GCN extracts crime tendency in communities that changes over time. The LSTM is used to extract trend of various crime over time across the Boston area. Then, a topology diagram of adjacent communities is constructed according to the neighborhood relations of the 12 partitions of Boston where the original data are located. Moreover, the spatial features extracted by GCN are stored in this diagram too. Finally, through the Self-Attention mechanism, the temporal and spatial features are integrated and the proportion of important features is increased in order to predict the number of crimes. In this way, a spatiotemporal prediction model is constructed successfully as a whole.

Taking Boston as an example, we collected e raw data about 310,000 crimes committed within three years in this place. Crime Data is open-source data collected from Kaggle [8], a website that provides data for competitions and new algorithms research. It offers free download service. The experimental results verify the validity of the data, which have a good prediction effect in a certain range and get a high prediction evaluation score. In addition, we conduct a comparative analysis between the experimental results with other algorithms proposed in similar studies. The results show that our model has a better performance in crime prediction.

The structure of the paper is as follows: The first part introduces the research background of this study. The second part is a literature review of crime prediction. The third section shows a general description of the Boston dataset. The fourth part introduces our prediction model. The fifth part introduces the research result and makes the related discussion. Finally, in the sixth part, a conclusion is drawn and directions are put forward for future research.

II. Related Work

The research report of American Rand Corporation suggests that predictive policing is a model that guides police departments to crack down on crimes through statistical analysis or predicts the most likely targets to prevent crimes [9]. United States is the first country who adopts data analysis to perform crime prediction. However, as the relevant factors of crime become more and more sophisticated, it is no longer possible to predict crime rate accurately by data analysis alone. Thanks to extensive application of artificial intelligence, crime forecasting can be more specific and precise. At present, crime prediction study focuses on exploring more correlative factors and improve the deep learning mode's performance to achieve a superior crime prediction system.

There is a long list of factors that induce crime committing.[10]. Recent researches focus on two dimensions: time and space when investigating relevant factors. In terms of time, holidays, seasons, and the weather are proved to be strongly associated with crime. Larceny is on an obvious rising trend during holidays since people flow on scenic

spots and shopping malls increased dramatically. Seasonal factors are strongly tied with crime rate. There is a significant increase in crime in summer as study proved. People' mood soured as summer comes, so it is harder for them to keep a sober mind. Weather conditions also play an important role. It is easier to commit a crime in sunny weather than on a rainy or snowy day. In terms of space, criminal types varied from region to region. The amount of crime in developed areas is significantly higher than that in underdeveloped areas. However, what is noteworthy is that drug-making crimes is an exception, which often took place in remote areas. At the same time, interaction effect exists in adjacent areas, which is similar to that of propagation in thermodynamics. Crime will also spread in space. The increase of crime amount in one area may influence the surroundings. Therefore, the lag of spatial crime is a weak link. In order to overcome this defect, we have tested a lot of deep learning models in our research. Finally, we extracted and fused spatiotemporal features successfully. Deep learning has the advantages of stability and scalability. The application of machine learning and deep learning model to fulfill predictive tasks has been a common practice in many fields, as well as crime prediction study. The model for crime prediction should be updated so that it could handle spatiotemporal features. In 2012, with the help of Markov chain, Sheng et al. predicted the station's annual maximum water level [11]. In 2014, R. Blair et al. adopted random forests to predict local violent crime [12]. In 2018, A. Torki et al. employed a mixture of firefly optimization techniques and fuzzy cognitive mapping neural networks to predict criminal activity [13]. In 2019, J. Hegde et al. used machine learning to evaluate engineering risk [14]. The usage of machine learning for predictive tasks is popular. In 2017, Bao W et al. pioneered the use of deep learning for real time crime prediction, using a spatiotemporal deep learning model in their paper [15]. In 2020, Han X et al. predicted the risk of theft in urban communities by using the model of combining LSTM with ST-GCN, which proved that different deep learning models could be used together, and making full use of time and space features can significantly improve the efficiency of crime prediction [16].

Attention mechanism, a kind of deep learning model, is widely used around the world. It is essentially like human visual attention mechanism. It can extract the key feature from all kinds of information. It includes soft attention mechanism and self-attention mechanism, both of which can be used to perform various predictive tasks. In 2017, the LSTM frame constructed by Fernando and others has been successfully used to predict human motion trajectory and detect abnormal events [17]. In 2017, the LSTM framework constructed by Baziotis and others has been successfully applied to the analysis of Message-level and Topic-based sensation by using attention mechanism [18]. So, it is feasible to use the attention mechanism and other deep learning models to realize the predictive task, and it could get job done better.

Previous studies provide a direction for in depth study on crime prediction. It is safe to say that an approach with time and space factors combined demonstrates superior performance in crime prediction. But no one has yet taken advantage of the attention mechanism to combine spatiotemporal features and increase the proportion of important features to achieve better predictive accuracy.

In this context, based on deep learning, this paper proposes a comprehensive model based on attention mechanism that combines LSTM with GCN to predict crime, and carries out temporal and spatial crime prediction for 12 communities in Boston.

III. Data

3.1 Data set description

Model we used for training data and making prediction was downloaded from Kaggle, which provided open source related to Boston crime data [19]. The dataset in fig. 1 actually recorded 310,000 crimes of all types in Boston between June 2015 and May 2018. As shown in the chart, the daily tendency of crime quantity presents a roughly inverted "V" shape, which may be caused by the seasonal change [20]. Therefore, we input "month" into the model as the influencing factor, because it could be key to explaining seasonal tendency in crime quantity. At the same time, we also found that the days are usually weekends or holidays when crime amount was much higher than that

at other times, as shows in the fig. 1. Hence, we chose holidays and days of the week as influencing factors. In spatial dimension, we select the longitude, latitude, and partition of the original data in order to determine the exact location of the crime and mapped it on the map. The attributes, description, and form of the data are shown in the table. 1.
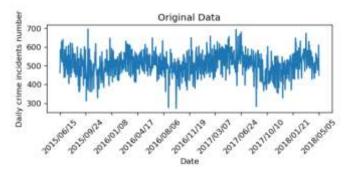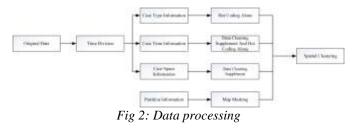

Fig 1: The daily number of crimes from June 15, 2015, to May 5, 2018.

Table 1 Feature Selection

Attribution	Description	Data Form
Case Type	OFFENSE_CODE_GROUP	String
Date Time	OCCURRED_ON_DATE	Date
Month	Month	Int
Week	DAY_OF_WEEK	String
Holiday	WHETHER THE HOLIDAY	String
Latitude	LAT	Float
Longitude	LONG	Float
Partition	DISTRICT	String

3.2 Data processing

First, the time data of every case needs to be divided and combined since the time information in each case of raw data is accurate to every minute and the case prediction is specific to a certain day or a certain week. Secondly, given the spatiotemporal heterogeneity indifferent types of cases, we need to conduce classification prediction on cases. It is required to carry out unique thermal coding of case categories in advance. Single-hot coding, also known as one-bit efficient coding, uses n-bit status registers to code N states. Each state has its own independent register bits, and at any time, only one of them is effective [21]. In addition, there are repeated alarms in the original data, and the space-time information of the case is missing, so it is necessary to repeat and complete the space-time information of the original data. Finally, the spatial information of the case is a kind of precise longitude and latitude coordinates, and the geographical range studied refers to surface information with a certain longitude and latitude interval. Hence, there is a need to map each case to the corresponding geographic grid block through spatial clustering [22]. The specific process is as follows: In the original data, there were 25,935 larceny cases, 16,548 drug cases and 23,633 assault cases. We treat case category and the statistics days as variables. We first carry out time data processing and date the incident to a specific day. The date of the crime is coded from Monday to Sunday as 0-6. After this, we carry out data cleaning, check weight, complete and other related work. This concludes time data processing. As for spatial data processing, we code the partition as 0-11, because the area where the case occurred has a total of 12 partitions in the data and is represented by letters and numbers.

Fig 2: Data processing

IV. Thodology

This section describes the proposed model framework and experimental details.

4.1 Proposed model framework

It can be seen from Fig. 3 that the integrated model we proposed can be categorized into 3 modules, namely: Spatio-temporal feature extraction module, time feature extraction module and attention mechanism integration module from bottom to top. Firstly, the spatial and temporal feature extraction module uses the GCN model to extract the spatial variation of crime incidents over time. Then, based on bidirectional LSTM, the temporal feature representation of the hidden layer output obtained by Attention is presented, and finally the classification is made by the full connection layer.
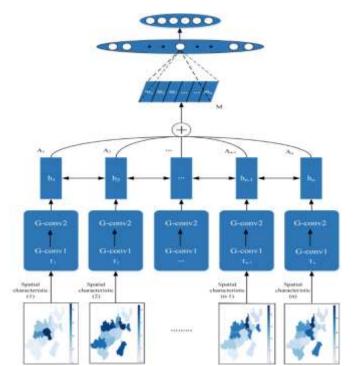

Fig 3: The hybrid GCN and LSTM structure based on attention mechanism.

In the integration model, each partition is set as a graph node that contains all the features of the crime data. GCN is used to process the graph data and LSTM is used to obtain the temporal characteristics of nodes.

4.1.1 Extracting spatio-temporal features
First proposed in Han X's paper on the risk prediction of urban community burglary crime, the fusion model of LSTM and ST-GCN can be used to predict crime data with spatial and temporal characteristics [23]. Its relevant definitions can be expressed as follows:

$$f(X, M) = \text{ReLU}(\widehat{D}^{-\frac{1}{2}}\widehat{A}\widehat{D}^{-\frac{1}{2}}XW + B)$$ (1)

Where X is the eigenmatrix, M is the adjacency matrix with self-ring added, D is its degree matrix, ReLU is the activation function of the network, and W and B are the network parameters. This formula is used to process graph data and extract spatiotemporal features.

4.1.2 Time feature extraction module

We employ LSTM model to build a time series model for the data processed by GCN to predict the tendency. LSTM consists of four gate functions and a memory cell namely, input gate, input modulation gate, forgetfulness gate, output gate and memory cell. LSTM can determine whether the new input is stored, forgotten, or as a unit of output is stored in the memory. It solves the problem of the short- and long-term dependence on the temporal crime data. So, the information in historical crime will never be forgotten. In our model, forward propagation calculation, backpropagation calculation and Adam parameter optimization algorithm of LSTM are used, and the learning rate is automatically adjusted by the activation function of ReLU. The gradient has an initial weight of 0.001. Genetic regression algorithm was used to adjust the learning rate. After 300 times of epochs, the no-loss value was reduced to 0.031 and then stopped changing, and the training ended.

4.1.3 Self-attention module

Self-attention was first introduced in Lin Z et al., aiming to solve the problem of low learning efficiency due to network deepening, which shows no effect on accuracy improvement or even reduce it [24]. Fig. 4 shows the details of the Self-attention mechanism. Supposing there are m cases, each case contains two dimensions, namely, time and space. By the embedding of each case to showcase the space-time characteristics of two-way LSTM output of hidden layers. We use a matrix to show it, where m represents the number of feature groups and b represents the length of the output vector of the hidden layer. Attention weight is calculated through all the following equations:

$$a = \text{softmax}(m_{s2}\tanh(M_{s1}H^T))$$ (2)

$$A = \text{softmax}(M_{s2}\tanh(M_{s1}H^T))$$ (3)

Where M_{s1} is the parameter matrix of $d_a * 2b$, and m_{s2} is the parameter vector of size d_a. After getting the weight, the weighted average of the hidden layer is carried out to get the expression of the number of output crimes. Embedding in each case focuses on some characteristics of each case, but there may be a combination of functions in each case to represent the complete characteristics of each case. For example, holidays, longitude and latitude is shown to significantly affect on the case, so more attention should be given to such characteristics. In order to explore more Attention, we expand m_{s2} to M_{s2} $(r * d_a)$,where r represents different characteristics drawn from the sentence. Then, we can get Attention matrix A. Finally, the embedding of the case is as follows:
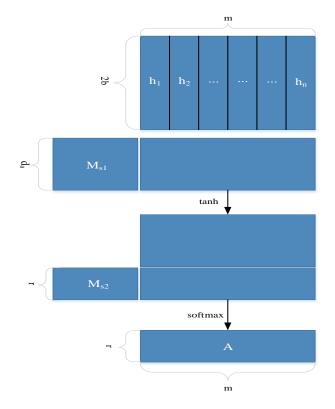
$$M = AH$$ (4)

Fig 4: The detailed diagram of self-attention model.

We adopt three indicators to evaluate the performance of the forecasting model, namely regression coefficient (R^2), root mean square error (RMSE), percent root means square error (PRMSE). The smaller RMSE and PRMSE, the higher of the prediction accuracy of the model. The closer R^2 gets to 1, the higher the accuracy [25]. They are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{5}$$

$$PRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}{\frac{1}{n}\sum_{i=1}^{n}y_i} \tag{6}$$

Where $\hat{y}_i$ and y_i are represent predicted value and true value respectively, n is the number of predicted samples.

Fig 5: The blue line represents the actual number of crimes, while the orange line represents the predictive number. In the first column, the crime type is Larceny, and the statistical days are 1, 4 and 7 from top to bottom for prediction. In the second column, the statistical period is 7 days, and the crime type from top to bottom is Drug Violation and the Cumulative of the three types of crime.

V. Results and Discussion

In this section, we will analyze and discuss the training results of proposed model and then evaluate the performance of the model based on different statistical days and different crime types. The training is set from June 15, 2015, to April 15, 2018, and the test is set from April 15, 2018, to June 29, 2018. The prediction is conducted under two dimensions: time and space [26].

Fig. 5 shows the effect of temporal dimension crime prediction, whose data was recorded every one day, four days, and seven days respectively in the first column. Through research, we have found that the crime prediction on a daily basis made little sense to basic-level police. They rely on their own experience and past data to predict crime in coming days, whose results shows no significant difference from our experimental result. However, as for leadership, predicting the number of crimes per week is quite beneficial for police dispatchment and alert status arrangement timely. Besides, work efficiency could be optimized successfully when prediction system established [27]. For this reason, we compared the prediction effects of three different statistical days. The PRMSE were 0.108, 0.085, and 0.066 in turn, and the R2 was 0.84, 0.76, and 0.83 in turn. When the statistic period was 7 days, the model has a relatively better prediction effect clearly. In other words, the model can be a great help for the forecast of the weekly alert situation. In order to verify the general applicability of the model, the second column of Figure 1 shows the crime types of Drug Violation, Assault, and cases collection. The cases collection means the total number of all kinds of crime, namely Drug Violation, Larceny and Assault. The PRMSE was 0.089, 0.078, and 0.061 respectively. In it, we merged simple assault and aggravated assault into assault, because we believe that neither simple assault nor aggravated assault can be ignored, and both result in personal injury

Fig. 6 shows the spatial map of average number of different crimes on weekly basis. We selected twelve communities in Boston that the data source was involved in. As can be seen from the figure, larceny is concentrated in the central and northwestern regions. According to the actual situation, we conclude that frequent occurrence of larceny is a strongly associated with locational factor. There are many large shopping malls and hospitals in city centers [28]. The maximum AE for each community was 5.4, the minimum AE was 0.3, and the average AE was 2.1. The number of drug-related crimes in the southeast and northwest regions is relatively high, which may be related to the convenience of smuggling in the coastal areas. The maximum AE for each community was 4.3, the minimum AE was 1.3, and the average AE was 2.7. Assault case also shows a high incidence in the northwest, with a maximum AE of 4.2, a minimum AE of 1.2, and an average AE of 2.5. Finally, the cumulative maximum AE of the three types of cases is 8.2, the minimum AE is 1, and the average AE is 4.1. In a sense, the country's tolerance of larceny and assault cases is greater than that of drug cases. However, in our model, the prediction error of drug violation crimes is the largest. Therefore, the prediction precision of our model still needs to be improved.
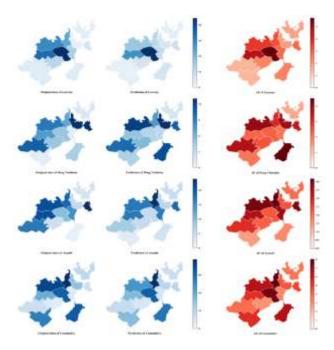
Fig 6:

The spatial distribution of original data, predicted data and Absolute Error (AE) of different communities in Boston. Original data is shown in the left column, predicted data in the middle column, and AE in the right column. In the first three rows, the crime types are Larceny, Drug Violation and Assault. The bottom is the cumulative of the three types. And the map quantity is the average number of crimes per day.

In addition, we compared the prediction accuracy of different deep learning models. The characteristics selected by LSTM are 'year', 'month', 'week', 'holiday or not'. The characteristic of our model is as same as LSTM and GCN, that is LSTM plus latitude and longitude processing in space. The crime type was set as theft, and the statistical days were set as every day. As can be seen from Table 2, our model has the largest R2, up to 0.84, which is a big leap compared to 0.31 in LSTM and a small progress compared to 0.80 in LSTM and GCN. At the same time, we get the minimum RMSE and PRMSE, that is 2.30 and 0.11 respectively. These two indicators are much lower than that in RMSE and PRMSE and slightly lower than that in LSTM and GCN.

Table 2 Comparison with Different Model

Model	R^2	RMSE	PRMSE
LSTM	0.31	4.90	0.22
LSTM And GCN	0.80*	2.41	0.13
Our Model	0.84*	2.30	0.11

*Denotes a significance level lower than 0.05

VI. Conclusion

Relatively speaking, crime predication is a complicated interdisciplinary research. Apart from geographic and time factors, advanced deep learning and machine learning models should be taken into account for better prediction accuracy. Abovementioned research must make itself with police forecasting in practice. The integrative model, LSTM and GCN, combines itself with Self-Attention mechanism proposed in this paper proved to be significant advantage in crime predication in the meantime, however, it faces some common problems that need to be addressed in future study.

Through the spatio-temporal comparative study of predicted crime quantity and actuals, as well as the superiority of proposed model over previous deep learning models, the higher prediction accuracy of our model is verified. Our model manages to combine the temporal characteristics in Boston with the spatial characteristics of communities. By doing so, we can predict crime tendency and accurate crime quantity effectively. This model has an advantage that LSTM and GCN convolution could process temporal characteristics simultaneously. Besides, Self-Attention mechanism in this model raises proportion of key feature effectively. Given the above, the model can provide a good reference for urban crime risk prevention.

However, problems such as data deficiencies or poor interpretability of this model need to be point out. First and foremost, the entered data in our model is relatively incomplete. Data we collected should be both dynamic and extensive as required under big data circumstances since crime committing is affected by social factors and weather factors besides seasons and holidays. So, more factors should be explored in future investigation. In addition, influenced by the structural complexity, many operational errors in this model need to be overcome in training process. The poor model interpretability is a common fault in machine learning algorithms. Although machine learning algorithms has been widely used in crime prediction, model's interpretability has decreased significantly. The interpretability plays a pivotal role in both model optimization and application [29]. Knowledge graph is characterized by structure and visualization. With the development of knowledge graph, there will be a new study focus on improving model's interpretability. Under the premise of ensuring high interpretability, it can be a new research direction that combining advantages of machine learning model with knowledge graph to improve the prediction effect continuously in the future [30].

References

[1] M. L. J. B. Mckinney, "Urbanization, Biodiversity, and Conservation," vol. 52, no. 10, pp. 883-890, 2002.

[2] Jihong et al., "Community policing: is it changing the basic functions of policing?: Findings from a longitudinal study of 200+ municipal police agencies," 2001.

[3] S. B. Mehta, R. D. J. I. J. o. E. Doshi, and T. Research, "Automatic Clustering Crime Region Prediction Model using Statistical Method in Data Mining," vol. V9, no. 4, 2020.

[4] P. J. J. E. o. G. Zak, "Larceny," vol. 1, no. 2, pp. 157-179, 2000.

[5] D. E. Joranson, A. M. J. J. o. P. Gilson, and S. Management, "Drug Crime Is a Source of Abused Pain Medications in the United States," vol. 30, no. 4, pp. 299-301, 2005.

[6] S. Feldman-Summers, K. J. C. J. Lindner, and Behavior, "Perceptions of Victims and Defendants in Criminal Assault Cases," vol. 3, no. 2, pp. 135-150, 1976.

[7] T. Almanie, R. Mirza, and E. J. C. S. Lor, "Crime Prediction Based On Crime Types And Using Spatial And Temporal Criminal Hotspots," vol. 5, no. 4, 2015.

[8] A. Narayanan, E. Shi, and B. I. P. Rubinstein, "Link Prediction by De-anonymization: How We Won the Kaggle Social Network Challenge," 2011.

[9] W. L. Perry, B. Mcinnis, C. C. Price, S. Smith, and J. S. Hollywood, The Role of Crime Forecasting in Law Enforcement Operations. RAND Corporation, 2013.

[10] B. Snook, R. M. Cullen, A. Mokros, S. J. J. o. I. P. Harbort, and O. Profiling, "Serial Murderers' Spatial Decisions: Factors that Influence Crime Location Choice," vol. 2, no. 3, pp. 147-164, 2005.

[11] Sheng et al., "The application of a Grey Markov Model to forecasting annual maximum water levels at hydrological stations," 2012.

[12] R. Blair, C. Blattman, and A. J. S. S. E. P. Hartman, "Predicting Local Violence," 2014.(suijisenlin)

[13] A. Torki, A. J. N. C. Mohammed, and Applications, "Crime activities prediction using hybridization of firefly optimization technique and fuzzy cognitive map neural networks," 2018.

[14] J. Hegde and B. J. S. S. Rokseth, "Applications of machine learning methods for engineering risk assessment – A review," vol. 122, 2019.

[15] B. Wang, D. Zhang, D. Zhang, P. J. Brantingham, and A. L. Bertozzi, "Deep Learning for Real Time Crime Forecasting," 2017.

[16] X. Han, X. Hu, H. Wu, B. Shen, and J. J. I. A. Wu, "Risk Prediction of Theft Crimes in Urban Communities: An Integrated Model of LSTM and ST-GCN," vol. 8, pp. 217222-217230, 2020.

[17] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Soft + Hardwired Attention: An LSTM Framework for Human Trajectory Prediction and Abnormal Event Detection," 2017.

[18] C. Baziotis, N. Pelekis, and C. Doulkeridis, "DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis," in Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 2017.

[19] https://www.kaggle.com/AnalyzeBoston/crimes-in-boston/download

[20] D. Examiner, "Police Prepare for 'Silly Season' Crime."

[21] B. Catalinotto, "System and method for establishing a local page and grid numbering system in a geographic referencing system," 2009.

[22] K.T. Cheng, S. J. C.-A. D. o. I. C. Devadas, and I. T. o. Systems, "Delay-fault test generation and synthesis for testability under a standard scan design methodology," vol. 12, no. 8, pp. 1217-1231, 1993.

[23] X. Han, X. Hu, H. Wu, B. Shen, and J. J. I. A. Wu, "Risk Prediction of Theft Crimes in Urban Communities: An Integrated Model of LSTM and ST-GCN," vol. 8, pp. 217222-217230, 2020.

[24] Z. Lin et al., "A Structured Self-attentive Sentence Embedding," 2017.

[25] K. E. Castellano and D. F. J. J. o. E. M. Mccaffrey, "Estimating the Accuracy of Relative Growth Measures Using Empirical Data," no. 4, 2020.

[26] F. Ilhan, S. F. Tekin, and B. Aksoy, "Spatio-Temporal Crime Prediction with Temporally Hierarchical Convolutional Neural Networks," in 2020 28th Signal Processing and Communications Applications Conference (SIU), 2020.

[27] A. G. J. S. S. E. P. Ferguson, "Predictive Policing and Reasonable Suspicion," vol. 62, 2013.

[28] D. Coronado and C. J. A. N. Luebbering, "Boston: Geography in a Sanctuary City," 2017.

[29] A. Jha, J. K. Aicher, M. R. Gazzara, D. Singh, and Y. J. G. B. Barash, "Enhanced Integrated Gradients: improving interpretability of deep learning models using splicing codes as a case study," vol. 21, no. 1, 2020.

[30] N. Qazi and B. L. W. Wong, "Behavioural & Tempo-Spatial Knowledge Graph for Crime Matching through Graph Theory," in 2017 European Intelligence and Security Informatics Conference (EISIC), 2017.