IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 26, 2020, accepted November 21, 2020, date of publication December 2, 2020,
date of current version December 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3041924

Risk Prediction of Theft Crimes in Urban Communities: An Integrated Model of LSTM and ST-GCN

XINGE HAN[1], XIAOFENG HU[1], HUANGGANG WU[2], BING SHEN[1], AND JIANSONG WU[3]
[1]School of Information Network Security, People's Public Security University of China, Beijing 102628, China
[2]School of International Police Studies, People's Public Security University of China, Beijing 102628, China
[3]School of Emergency Management and Safety Engineering, China University of Mining and Technology, Beijing 100083, China

Corresponding author: Xiaofeng Hu (huxiaofeng@ppsuc.edu.cn)

This work was supported in part by the National Key Research and Development Project of China under Grant 2018YFC0809700, and in part by the Ministry of Public Security's Special Project for Strengthening Police Science and Technology Infrastructure of China under Grant 2018GABJC01.

ABSTRACT Urbanization has been speeding up social and economic transformations in urban communities, the smallest social units in a city. However, urbanization brings challenges to urban management and security. Therefore, a system of risk prediction of crimes may be essential to crime prevention and control in urban communities and its system improvement. To tackle crime-related problems in urban communities, this paper proposes a model of daily crime prediction by combining Long Short-Term Memory Network (LSTM) and Spatial-Temporal Graph Convolutional Network (ST-GCN) to automatically and effectively detect the high-risk areas in a city. Topological maps of urban communities carry the dataset in the model, which mainly includes two modules — spatial-temporal features extraction module and temporal feature extraction module — to extract the factors of theft crimes collectively. We have performed the experimental evaluation of the existing crime data from Chicago, America. The results show that the integrated model demonstrates positive performance in predicting the number of crimes within the sliding time range.

INDEX TERMS Crime prediction, crime rates, graph convolutional network, long short-term memory network, spatial-temporal.

I. INTRODUCTION

Nowadays, the majority of the world's population lives in urban areas, and the proportion continues increasing as people move to fast-developing cities to fulfil their needs and aspirations [1]. Meanwhile, social, economic, and environmental threats to urban areas have emerged as adverse effects of urbanization. For example, it presents challenges to the organizations responsible for city management and essential service provision, like resource planning (water and electricity), transit, air and water qualities, and public safety [2]. Moreover, for the cities with higher crime rates, crime spiking is becoming one of the most critical public issues, threatening social stability and economic progress in different ways [3], [4]. As the smallest constituent units of a city, urban communities meet the daily needs of urban residents for living, work, leisure, and so forth. As science and technology develop, crime prevention and control systems for urban communities are also improving [5]. Risk prediction of crimes in urban communities can effectively improve the overall level of crime prevention and control in urban areas.

Theft is one type of common property crimes all over the world, which generally refers to the act of illegal possession of the public or other people's properties. In recent years, despite generally downward crime rates, preventing and combating crime remains a challenge. This may be explained by the following facts: more and more criminals have been equipped with high-tech tools and even well trained; local governments and the police often pay more attention to murder, assault, and other violent crimes [6], so limited resources are allocated to prevent or fight against property crimes. Thus, the rate of theft crimes seems much higher than that of violent crimes. Moreover, effective prevention strategies against new types of theft (such as theft of electric vehicles) are quite

The associate editor coordinating the review of this manuscript and approving it for publication was Bohui Wang.

inadequate. Quantitative and accurate risk analysis is vital to prevent theft when police resources are limited. Besides, there are strong correlations between crime rates and other variables such as the geographic location of a community (with low-risk and high-risk areas), time and climate during the year (seasonal patterns) [7]. An accurate prediction model of crime in urban communities is required to detect the communities more vulnerable to crime incidents. This method can optimize the allocation of police resources, thereby deploying police officers to high-risk areas or dismissing police force from the areas with declining crime rates to prevent or quickly respond to criminal activities.

In order to analyze both spatial and temporal feature extraction trends of crime to automatically detect the high-risk communities in urban areas and predict the number of crimes in each community, this paper proposes a new model based on LSTM and ST-GCN. The algorithm includes three steps. First, a topological map of neighboring communities is made according to the regional location and neighboring relationships of urban communities, and each node on the topological map stores historical crime data, the weather data and the holiday data of the community. Second, ST-GCN is utilized to capture the transition tendency of crimes between neighboring communities over time (spatial-temporal features extraction module); LSTM is used to capture the trend of crimes in each community over time (temporal feature extraction module). Third, a Gradient Boost Decision Tree (GBDT) is used to integrate the predicted values from the spatial-temporal features extraction module and the ones from the temporal feature extraction module. The result is a spatial-temporal model of crime predicting, which consists of a set of topological maps of neighboring communities and a set of crime predicting modules to predict the number of crimes in each community.

Taking Chicago as a case study, we made an analysis of about 0.32 million crimes within a period of six-year or so. The crime data were collected from Chicago's open data portal [8], a web framework that provides data, map and graphs about the city, and free download service. The experimental evaluation result substantiates the effectiveness of the approach by achieving good evaluation scores in the spatial-temporal prediction of the number of crimes in the sliding time range. In addition, we conducted a comparative analysis of the results with other algorithms presented in previous studies and got a higher evaluation score of the former as opposed to the latter.

The paper is organized as follows: Section 1 is an introduction to the research background. Section 2 is a literature review of crime data mining. Section 3 shows a general description of the Chicago data set, in particular the theft data. Section 4 focuses on the spatial-temporal crime prediction algorithm, describing its steps in detail and the experimental evaluation of a case study. The experimental results are presented and discussed in Section 5. Finally, Section 6 draws a conclusion and points to the future research.

II. RELATED WORK

Crime prediction is a scientific methodology to analyze, excavate and investigate existing crime dataset, information and other variables positively correlated with crimes. It analyzes the number and tendency of crimes that may occur in specific spatial-temporal nodes in the future, as well as focuses on prediction and inference of the probability of crime or re-offending [9]. In the United States, data analysis was first proposed and applied to crime prediction [10]. However, due to the high-dimensional nature of crime factors, traditional crime analysis that only rely on human resources cannot adequately predict crimes. With the development of intelligent technology, such as machine learning and deep learning, crime prediction has become a research focus [11].

In terms of prediction content, the prediction of crimes is roughly subcategorized as follows: re-offending prediction [11], [12], victim prediction [13], offender prediction [14], [15] crime pattern prediction [16], [17] and crime hot spots prediction [18]. Crime hot spots prediction can be classified into temporal prediction [19], spatial prediction [20] and spatial-temporal prediction [21], [22].

The spatial-temporal prediction of crime hot spots relies on the theory of near repeat of crime [23], which means the risk of a certain type of crime may spread in a particular time-space range, showing specific spatial-temporal correlations between new crime data and historical crime data. However, this correlation gradually weakens with the expansion of the time-space range. When a crime occurs in a specific place, the occurrence probability of the same type of crime in the area adjacent to it will significantly increase in a short time. The theoretical backgrounds of the near-repeat theory are daily activity theory [24], rational choice theory [25], and foraging theory [26]. According to daily activity theory, the emergence of crimes is affected by three major factors — people with criminal motives, suitable criminal targets, and the lack of guardians. When the three elements occur, the possibility of crime will show an upward trend. If the occurrence coincides in time and space, it may lead to crimes. The rational choice theory mainly interprets the impact of its action purpose on the possibility of crime from the perspective of an attacker. The theory explains that under the premise of rational choice when there are different action strategies to choose from in a particular environment, individuals subjectively have different preference arrangements for the results caused by different choices. Actors often decide to minimize the cost but maximize the benefits by an action strategy, that is, rational selectors tend to choose an optimal strategy and maximize the benefits at the minimum cost. Foraging theory was introduced from behavioral ecology and was initially used to predict the possible infested locations of carnivorous animals by analyzing their foraging patterns. It treats the offenders as a kind of "foraging" animal to analyze their possible crime locations and crime patterns. The theory holds that there is a specific spatial pattern between the offender's foothold and the crime location, and the offender often chooses a place he is familiar with to commit a crime. The crime continues in and around the area [27]. The above theories provide essential theoretical bases for the near-repeat theory of crime. It is worth noting that an actor in the case must be a sufficiently rational individual since near-repeat theory originates from rational choice theory. Therefore, crimes of passion between acquaintances are not within the scope of near-repeat theory.

The near-repeat phenomenon was first observed in burglary cases [28]. A scholar from the University of London in the United Kingdom studied the theft data in Merseyside, the UK in 2004, and found that the burglary rate, within a 400-meter area around the stolen household significantly increases within two months after a burglary case occurs. Since then, the near-repeat phenomena were noticed in more and more types of cases, including armed robbery [29], motor vehicle theft [30], shooting [31], and urban grenade explosion attacks [32] and so on.

In order to analyze the near-repeat phenomenon of crimes in the spatial-temporal range, the methods of predicting the space-time distribution of crimes are commonly used, for example, Zhang J *et al.* [33] proposed a method based on deep learning, called spatial-temporal residual network (ST-ResNet), to simultaneously predict the inflow and outflow of passengers in each area of a city. Wang B *et al.* [21] used ST-ResNet to predict crime distribution in Los Angeles. Since the above-mentioned methods only make predictions based on spatial-temporal grids and are inapplicable to topological spaces, Kipf TN *et al.* [34] proposed a Graph Convolutional Network (GCN) as an active variant of the convolutional neural network. Then, Yan S *et al.* [35] combined ST-ResNet and GCN to propose ST-GCN for human action recognition. This method can capture the transition of criminal events in a topological space to remedy the deficiency of traditional spatial-temporal prediction model. ST-GCN was quickly applied to motion recognition and traffic. For example, Kong Y *et al.* [36] built a dynamic skeleton model based on ST-GCN, combined with the attention module; Geng X *et al.* [37] proposed a spatial-temporal multigraph convolutional network based on ST-GCN (ST-MGCN) to forecast the demand for rides. ST-GCN has a good prediction effect on the transition of crimes in a topological space, but it attaches less importance to timing changes on a single node. For the crime field, the number of crimes in each community is influenced by the transition of crimes in surrounding communities and the previous number of crimes in this community. Therefore, it is difficult to predict the crime risk accurately in a single community only through ST-GCN.

Previous studies could provide references for crime prediction research, but few made spatial-temporal crime prediction analysis based on a topological space, or combined temporal prediction and spatial-temporal prediction to forecast community crime risk.

In this context, this paper, based on near-repeat theory and deep learning, proposes an integrated model of crime prediction by combining LSTM and ST-GCN to analyze the risk of crime in a specific time-space range in urban communities.

III. DATA

A. DATA SET DESCRIPTION

The data utilized to train the models and perform the experimental evaluation are housed on Chicago open data portal, a web framework developed (and currently managed) by the Chicago government that provides data, map and graphs about the city, and free download service. The crime data were collected from the 'Crimes - 2001 to present' dataset, a real-life collection of criminal cases happening in Chicago from 2001 to present. We selected 77 communities in Chicago, among which the largest and the smallest communities have areas of 371.8 square kilometers and 19.9 square kilometers, respectively. Starting from the 'Crimes - 2001 to present' dataset, we collected all criminal cases within the communities over six years (1985 days), from January 1, 2015, to March 10, 2020.

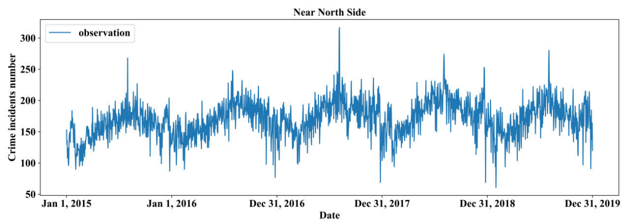


FIGURE 1. The daily number of crimes from January 1, 2015, to December 31, 2019.

Figure1 shows the daily number of crimes from January 1, 2015, to December 31, 2019. It is evident that the number of theft crimes in Chicago each year shows a steady reversed "V" pattern in which the occurrence of crimes varies with season. The number of crimes significantly increases in late spring, peaks during the summer, decreases in autumn and generally falls in winter. Through an analysis of the impact of climate change on crime [7], we argue that temperature variation may explain the seasonality. Therefore, the weather data are entered as an external environmental factor into the model for training.

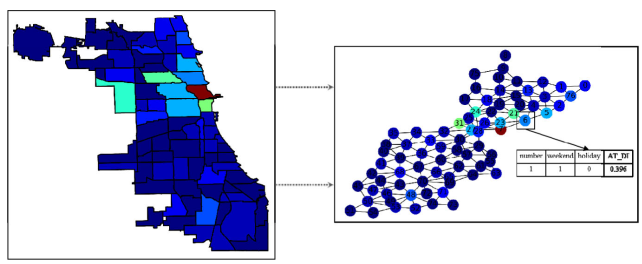


FIGURE 2. The topological map of adjacent communities and the node data. The feature 'number' indicates the number of crimes in a day; the feature 'holiday' indicates whether the day is a weekend or not, if it's a weekend, its value is 1, otherwise, its value is 0; the feature 'holiday' indicates whether the day is a holiday or not, if it's a holiday, its value is 1, otherwise, its value is 0; the feature 'AT_DI' means heat stress calculated by sensible temperature testing.

B. DATA PREPROCESSING

The topological graph of communities and the data placed in each graph node are shown in Figure 2 where the feature 'number' represents the number of crimes in communities, the feature 'weekend' represents whether the day is a weekend or a weekday, and the feature 'holiday' represents whether the day is a holiday or not. The feature 'AT_DI' is the thermal stress calculated based on temperature, humidity and wind speed, and the calculation formula is as follows:

$$\mathrm{AT\_DI} = 0.5T_W + 0.5AT \tag{1}$$

where AT is the apparent temperature (°C), and its value is approximately calculated by equation (2); T_W is the thermodynamic wet-bulb temperature (°C), and its value is approximated by equation (4);

$$AT = 1.07T + 0.2e - 0.65V - 2.7 \tag{2}$$

where AT is the apparent temperature (°C); T is the air temperature (°C); e is the water vapour pressure (hPa), and its value is approximately calculated by equation (3); V is the wind speed (m/sec);

$$\mathrm{e} = \frac{RH}{100} \times 6.105 \times \mathrm{e}^{\frac{17.2\mathrm{T}}{237.7+\mathrm{T}}} \tag{3}$$

where RH is relative humidity (%);

$$\begin{aligned} T_w = {} & AT \arctan(0.151977\sqrt{RH + 8.313659}) \\ & + \arctan(AT + RH) - \arctan(RH - 1.676331) \\ & + 0.00391838RH^{\frac{3}{2}} \arctan(0.023101RH) - 4.68035 \end{aligned} \tag{4}$$

A total of 145,992 spatial-temporal crime cases were gathered in the process, of which 145,679 crime cases collected from January 1, 2015, to December 31, 2019, are used as a train set, and 5,313 crime cases collected from January 1 to March 10, 2020, are used as a test set. The above-mentioned data have different roles in the model training process according to distinct time characters. For instance, the data before 2020 are only regarded as a temporal feature; the other data are used as both a temporal feature and a label. That will be explained in detail in the following section.

IV. METHODOLOGY

This section will describe the proposed model framework and experimental details.

A. PROPOSED MODEL FRAMEWORK

From Figure 3, we can see the proposed integrated model, which can be categorized into three modules—spatial-temporal feature extraction module, temporal feature extraction module, and feature integration module. First, the spatial-temporal feature extraction module is a combination of GCN and ST-ResNet (ST-GCN) to extract the transition of crimes in space over time. Then, in order to detect crimes in each community, the temporal feature extraction module is built based on the LSTM network. Finally, the feature integration module employs GBDT model to integrate the predicted values from the spatial-temporal feature extraction module and the temporal feature extraction module.

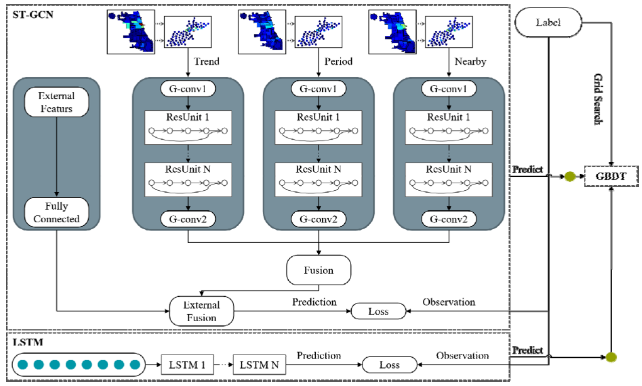


FIGURE 3. The structure of the integrated model, where blue points and green points refer to inputted temporal data and the predicted data, respectively.

1) SPATIAL-TEMPORAL FEATURES EXTRACTION MODULE

In this module, community is set as a graph node with the features included in the historical data of crimes. We employed Spatial-Temporal Graph Convolutional Networks (ST-GCN) to deal with the graph data. Graph Convolutional Networks (GCN) was used to deal with this graph, and Spatio-Temporal Residual Networks (ST-ResNet) was applied to extract the spatial-temporal features within each graph set.

GCN, based on Convolution Neural Network (CNN), was first proposed by TN Kipf *et al.* (2016) [34], which can be leveraged for the graph structure data in deep learning. Its definition is given as follows:

$$f(X, A) = \mathrm{ReLU}(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}X\mathrm{W} + \mathrm{b}) \tag{5}$$

where X is the feature matrix, A is the adjacency matrix with added self-loops, D is its degree matrix, ReLU is the activation function of the network, W and b are the parameters of the network.

As can be seen from the definition above, a GCN layer can get the effects of the nodes around each node and attach them to it. However, in view of topology with such a large number of nodes, the transition effects between them must be considered. In this sense, we need numerous GCN to capture further transition effects from long-range or even citywide dependencies, which requires a very high network depth. Besides, the transibility of crime is influenced by not only nearby feature, but also periodic feature and trend feature. Therefore, it is necessary to set ST-ResNet as a base model to host the GCN layer (ST-GCN).

ST-ResNet was proposed by J Zhang [33] based on deep Residual Network (ResNet) [38] to solve the problems of inefficient learning and inability to significantly improve accuracy or even reduced accuracy due to network deepening. On this basis, we define the nearby data as the data collected within three days before the date of prediction; the periodic data as the data collected within three weeks before the date of prediction; the trend data as the data collected within three years before the date of prediction.

Previous research clearly shows that there are strong correlations between crimes and complex external factors. In our experiment, weather, holiday and weekday fall within the external data (see Fig.2). Unlike other features, the future weather cannot be ex ante known, and forecasting weather will represent the weather in date d.

2) TEMPORAL FEATURE EXTRACTION MODULE

Crime cases are collected as the timing data when Long Short-Term Memory network (LSTM) [39] is working in this module to compute the number of crimes.

In order to process the timing data precisely, LSTM, an improved multilayer perceptron network based on Recurrent Neural Network (RNN), is used in this model. LSTM can learn the long short-term dependence information between the serial data by determining whether the new input is stored, forgotten, or stored in the memory unit as an output.

The latest output of the timing data is relevant to early input and output, that is, the output depends on the input and the early memory. As an improved version of RNN, LSTM also includes forward propagation calculation, backpropagation through time algorithm (BPTT) and Adam parameter optimization algorithm. The difference lies in that LSTM has made a certain transformation of the RNN's memory, screening the memory information, and only transmitting the information to be memorized. In this way, gradient disappearance or explosion caused by the growth of the dependent sequence and the increase of the multiplicative term during the backpropagation of the model may be prevented. Specifically, LSTM sets a gate to enable historical data of crimes to pass through selectively, thereby filtering or adding corresponding crime information to memory. LSTM adds historical data of crimes and the current inputted number of crimes so that previous memories will continue to exist instead of partially disappearing due to multiplication. Therefore, LSTM will not cause the attenuation of the effective information on historical crimes long ago and can deal with long-term memory problems.

The structure of one layer of LSTM is shown in Fig 4, which is the state of a cell at different moments. The four small yellow rectangles are the hidden layer structures of the ordinary neural network. The activation function of the third small yellow rectangle is tanh and the activation functions of the rest are sigmoid. The input X at time t and the output $h_{(t-1)}$ at time *t-1* are spliced and then inputted into the cell. Therefore, the input to LSTM includes not only the original data set but also the output of previous moment ($h_{(t-1)}$). The memory-related part is entirely controlled by various gate structures (0 and 1). The cell is roughly divided into two horizontal lines: the upper horizontal line controls long-term memory, while the lower horizontal line controls short-term memory.

B. CASE STUDY

This solution leverages an integrated architecture of LSTM and ST-GCN.

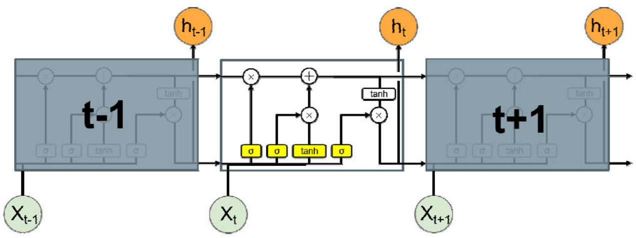


FIGURE 4. The structure of one layer of LSTM at time step $t-1$, t, and $t+1$.

In the LSTM model, the data are inputted into the model through a sliding window, the size of that is set as 8, that is, the data of eight days prior to the date of prediction for a particular community are used to predict the number of crimes in that community on that date. Only the timing data of the number of crimes are used for prediction. The input layer connects 5 LSTM layers and is activated by the ReLU activation function, and the output layer with one node is obtained. The model uses RMSE as the loss function, and uses the NAdam optimizer to optimize it. The learning rate and epoch are set, automatically adjusted by the model. The initial learning rate is 0.01. The learning rate is adjusted by using the Cosine Annealing method. When no loss value lowers after 100 times of epoch, the training will terminate.

In the operation of the ST-GCN model, firstly, the data are divided into three groups, namely, trend data, period data and nearby data. Each type of data passes through a graph convolution layer and 18 residual units, and each layer of residual units contains four graph convolution layers. Finally, the results of the three sets of data are fused. After being inputted, the external data pass through a fully connected layer and fuse with the graph convolution data in the external fusion layer. In the experiment, ReLU is used as the activation function, and the output layer with one node is finally obtained. The loss function and training method selected for this model consist with that of the LSTM model.

After the prediction of the two sets of models, the results enter into Gradient Boosting Decision Tree (GBDT) regressor for final fusion. The regressor uses a grid search method to adjust parameters, and the loss function is also RMSE.

Actually, the model in training only predicts one-day data at a time. After 69 times of training and predicting, 69 sets of prediction data are obtained for the final model verification.

Two metrics are used to evaluate the performance of the prediction model: Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), which are defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}} \sum_{i=1}^{n} (origenal_i - predicted_i)^2} \tag{6}$$

$$\mathrm{MAPE} = \sum_{i=1}^{\mathrm{n}} \left| \frac{origenal_i - predicted_i}{origenal_i} \right| \times \frac{100}{n} \tag{7}$$

where n is the total number of days.

V. RESULTS AND DISCUSSION

In this section, we will analyze and discuss the experimental results and evaluate the performance of the proposed model in each community each day within the test data set. The test set is set from January 1, 2020, to March 10, 2020. For both evaluation indexes, the model's prediction performance is the main concern in our discussion.

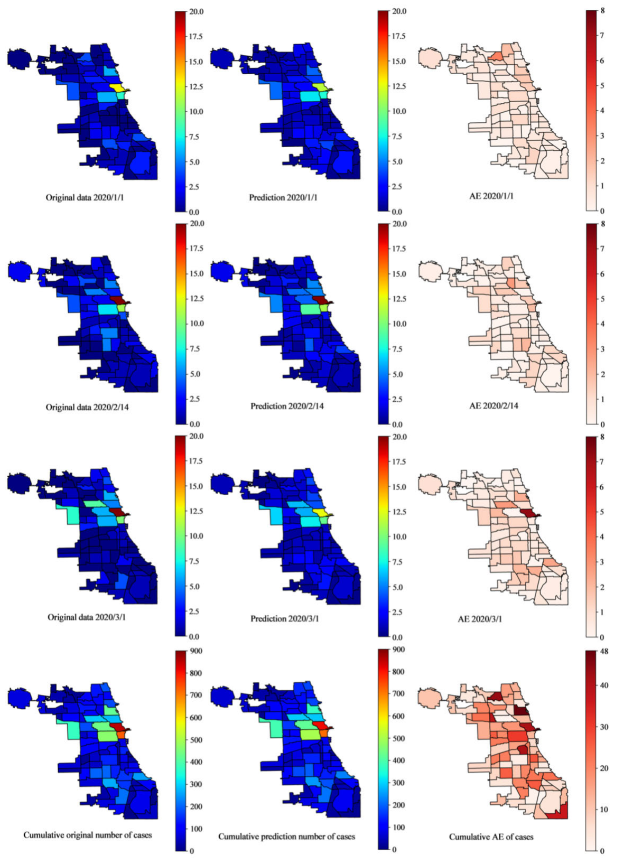


FIGURE 5. Spatial distribution of the predicted number of crimes and Absolute Error (AE) of different communities in Chicago. Left panels show the original crime data, middle panels show the prediction of the number of crimes, and right panels show the absolute error between prediction and observation. Bottom panels show the accumulations for the number of crimes and AE from January 1, 2020, to March 10, 2020.

Fig. 5 shows the distribution of actual incidents and Absolute Error (AE) within three days, and the accumulative incident quantity and AE within 69 days. On January 1, 2020, with the max AE of 3.4 (in the community of North Park), the average AE is 0.62; on February 14, 2020, a weekday, with the max AE of 2.7 (in the community of North Center), the average AE of 0.61; on March 1, 2020, a weekend, with the max AE of 7.2 (in the community of Near North Side), the average AE of 0.87; on the accumulative incident quantity, with the max AE of 48.39 (in the community of Lake View), the average AE of 14.62.

The four communities with the largest AE are all located in the northeast of Chicago. This area is adjacent to a major tourist attraction, Lake Michigan, which seems to be strongly associated with the local situation. The reasons for

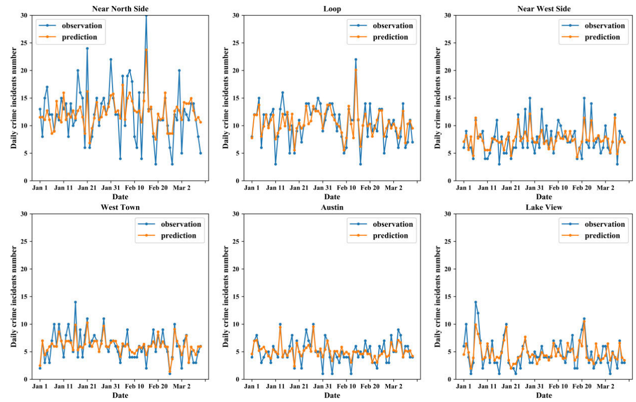


FIGURE 6. The daily crime prediction of six communities in Chicago. Blue lines and points represent the actual daily number of crimes, and orange lines and points represent the predictions.

the model's better prediction of weekdays than holidays and weekends are irregular life pace and the influx of tourists during the holidays. Therefore, it is inadequate to estimate the crime situation during the holidays only through the existing features, and more features (social factors) need to be taken into account.

However, when it comes to application, a priority must be given to the accurate prediction of the communities with a large number of crime cases. For this reason, we selected six communities within a specified period of time, and made a comparison between the daily average actual value and the predicted value in each community, as shown in Fig. 6. These six communities are Near North Side, with the fewest actual crime cases of 3, the most crime cases of 30, the maximum AE of 7.8, and the average AE of 2.67; Loop, with the fewest actual crime cases of 3, the most crime cases of 22, the maximum AE of 4.54, and the average AE of 1.38; Near West Side, with the fewest actual crime cases of 3, the most crime cases of 15, the maximum AE of 4.13, and the average AE of 1.33; West Town, with the fewest actual crime cases of 1, the most crime cases of 14, the maximum AE of 4.14, and the average AE of 0.99; Austin, with the fewest actual crime cases of 1, the most crime cases of 10, the maximum AE of 5.7, and the average AE of 0.94; Lake view, with the fewest actual crime cases of 1, the most crime cases of 14, the maximum AE of 4.09, and the average AE of 1.17.

It is obvious that the biggest fluctuation in terms of the number of crimes appears in Near North Side. Although the model's predicted value of the number of crimes in the community is not as accurate as that in the other five communities, it grasps the change tendency of the number of crimes. It can be concluded that the social structure of Near North Side is relatively complicated because an influx of tourists arrive here every day. Due to the fact that constant changes of social factors lead to the fluctuating number of theft crimes, more social factors need to be included in our model to capture the variation pattern of the number of crimes in the community.

In order to figure out whether this fitting effect exists in all the 77 communities, we average the daily number of crimes

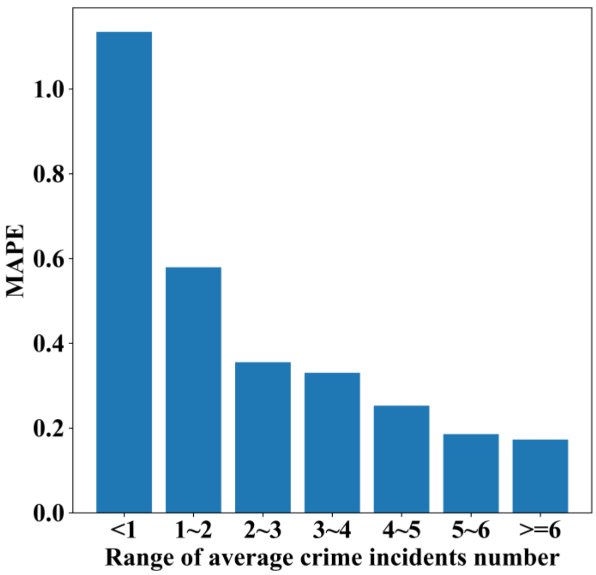


FIGURE 7. MAPE for the average number of crimes in a community. Blue bars represent the MAPEs in the communities with the average number of crimes no more than one, between one to two (including two), and so on.

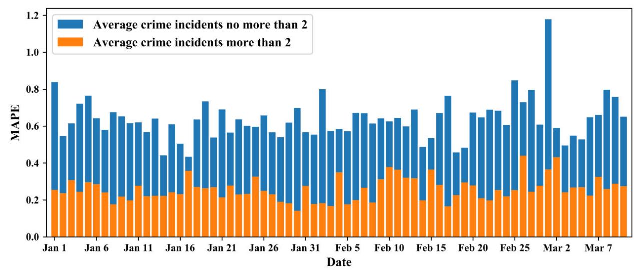


FIGURE 8. MAPE for the average number of crimes in a community. Blue bars and orange bars represent the MAPEs in the communities with the average number of crimes smaller than two and more than two, respectively.

in each community from January 1, 2020, to March 10, 2020. The MAPEs predicted by the model are calculated separately for different average number of crimes in the communities. The results are shown in Fig. 7. When the average number of crimes is no more than two, the MAPE is almost 0.6; when that number is more than two, the MAPE is lower than 0.4. It also shows that the larger the average number of crimes in the community, the lower the MAPE between the predicted value and the actual value, and the better the fitting effect.

For the communities with the average number of crime no more than two, the model's MAPE value is very high, which means that the model cannot capture the variation pattern of the number of crimes in such communities.

A comparison of the MAPEs of the predicted results for the communities with the average number of crimes smaller than two and those with the average number of crimes more than two is drawn in Fig. 8. In terms of MAPE, the former is about twice as much as the latter.

The predicted errors of these communities are analyzed to explore the differences in the MAPE results. Through an analysis, no matter it is a community with a large number

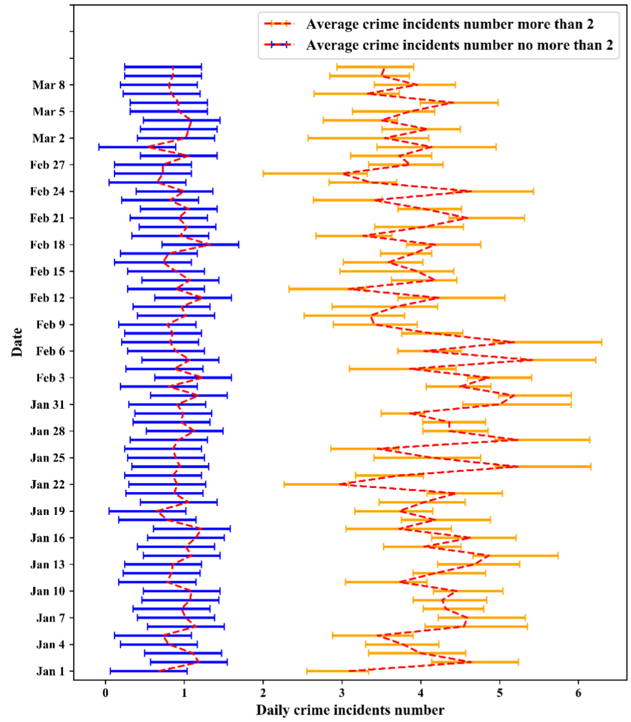


FIGURE 9. Average errors for the average number of crimes in different communities. The two red broken lines represent the average number of real daily crimes; blue error bars represent the average error between prediction and observation in the communities with the average number of crimes smaller than two, and orange error bars represent the average error between prediction and observation in the communities with the average number of crimes more than two.

of crimes or with a small number of crimes, the prediction error of the model is all about one, as demonstrated in error bars (Fig. 9), which is key to explaining the difference of the MAPE results in these two communities.

For the communities with a large number of crimes, the acceptable prediction error has less influence on the community. However, for the communities with a small number of crimes, this error margin cannot be accepted. Hence, the prediction accuracy of our model for the communities with no more than two crimes still needs to be improved. In application, it means that the community's security situation is tolerable if the number of crimes is no more than two within a single day because the occurrence of theft crimes in such a community tends to be probabilistic, but not affected by environmental factors. In a sense, our model may turn out to be effective, although it demonstrates poor predictability of crimes in the communities with a small number of crimes.

In addition, we compared the accuracy of our model with other traditional crime prediction models within the same data set by using the linear model, the tree model, and the timing model, as shown in Table 1. In this experiment, other models were trained to use the same data and features as our model (i.e., training without parameter adjustment or with simple parameter adjustment) to select the best performing linear model, tree model, and timing model. And then, Ridge, Random Forest, and LSTM were chosen as

TABLE 1. Comparison with different model.

Model	MAPE	RMSE	R^2
Ridge	0.54	4.04	0.48
Random Forest	0.51	3.91	0.53
LSTM	0.43	1.59	0.75*
Our Model	0.39	1.03	0.84*

* Denotes a significance level lower than 0.05.

comparison models, whose evaluation results (MAPE and RSME) in the initial training are higher than other similar models (ST-GCN is not used in comparison since it has fewer predictive results than LSTM). Besides, sliding window prediction was applied in this process (only one day's data are predicted at a time, and 69 times of training and prediction are completed simultaneously to obtain all the predictions from January 1, 2020, to March 10, 2020). Global tuning of Ridge and Random Forests are performed based on grid search. The training method of LSTM is the same as that of the LSTM part in the temporal feature extraction module (seen in 4.2.1). The communities with the number of crimes more than two are selected to calculate MAPE, RMSE and R^2 of their predictions.

The result indicates that our model has a lower value than the other models listed in Table 1 in terms of MAPE and RMSE, and the R^2 shows a higher value than the other models, meaning that its performance is better than the others. Moreover, the significance levels of the regression models are investigated. The results show that the significance levels of LSTM and the Integrated Model are both lower than 0.05, and that of ridge or Random Forest is higher than 0.05, which indicates that the regression results of LSTM and the Integrated Model we proposed are significant and reliable while those of ridge and Random Forest are not significant. Due to spatial factors, our model shows an improvement compared with the one that only uses LSTM for prediction.

VI. CONCLUSION

This paper investigates crime prediction through an integrative model of LSTM and ST-GCN by using the crime data and external social factors collected from the communities of Chicago.

A comparative study of the predicted results and actual crimes, as well as the superiority of the proposed model over traditional models, verifies that the proposed method is highly efficient for crime prediction in urban communities. The proposed method, by combining the temporal feature within a community and the spatial-temporal features between communities, is possible to capture the occurrence of crime and effectively predict the trend in crime, and then predict the number of theft crimes in communities the next day. The advantage of this model lies in that an integration of ST-GCN and LSTM helps to reasonably define the weighted relation between the impact of the number of historical crime incidents to be predicted in a community and the impact of the transition of crime incidents from surrounding communities in the prediction process. Besides, the temporal and spatial factors can be combined more effectively. As a result, the model captures crime patterns more accurately than the other models mentioned in this paper. The prediction results of the model may provide a valuable reference for crime risk prevention and control in urban communities.

However, there is still much research to do. Our experimental results indicate that the model is influenced too when the communities are severely affected by social variables. Therefore, apart from weather and time, social factors should be included in the future study. Besides, some training difficulties remain to be overcome since the structure of this model is relatively complicated.

The future study will focus on a more effective but precise model for crime prediction by adding more external features and simplify its structure.

REFERENCES

[1] U. Habitat, *State of the World's Cities 2012/2013: Prosperity of Cities*. Evanston, IL, USA: Routledge, 2013.

[2] F. Cicirelli, A. Guerrieri, G. Spezzano, and A. Vinci, "An edge-based platform for dynamic smart city applications," *Future Gener. Comput. Syst.*, vol. 76, pp. 106–118, Nov. 2017.

[3] M. A. Tayebi, M. Ester, U. Glässer, and P. L. Brantingham, "CRIMETRACER: Activity space based crime location prediction," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2014, pp. 472–480.

[4] H. Wang, D. Kifer, C. Graif, and Z. Li, "Crime rate inference with big data," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 635–644.

[5] A. Crawford and K. Evans, "Crime prevention and community safety," Oxford Univ. Press, Oxford, U.K., Tech. Rep. 35, 2017.

[6] M. Friedman, A. C. Grawert, and J. Cullen, "Crime trends, 1990–2016," Brennan Center Justice, New York Univ. School Law, New York, NY, USA, Tech. Rep., 2017.

[7] X. Hu, J. Wu, P. Chen, T. Sun, and D. Li, "Impact of climate variability and change on crime rates in tangshan, China," *Sci. Total Environ.*, vol. 609, pp. 1041–1048, Dec. 2017.

[8] M. Kassen, "A promising phenomenon of open data: A case study of the chicago open data project," *Government Inf. Quart.*, vol. 30, no. 4, pp. 508–513, Oct. 2013.

[9] P. Homel and N. Masson, "Partnerships for urban safety in fragile contexts: The intersection of community crime prevention and security sector reform," Geneva Peacebuilding Platform, Geneva, Switzerland, White Paper 17, 2016.

[10] E. W. Burgess, "The study of the delinquent as a person," *Amer. J. Sociol.*, vol. 28, no. 6, pp. 657–680, May 1923.

[11] D. M. Gottfredson, "Assessment and prediction methods in crime and delinquency," in *Contemporary Masters in Criminology*. Boston, MA, USA: Springer, 1995, pp. 337–372.

[12] J. E. Johndrow and K. Lum, "An algorithm for removing sensitive information: Application to race-independent recidivism prediction," *Ann. Appl. Statist.*, vol. 13, no. 1, pp. 189–220, Mar. 2019.

[13] Y. Boshmaf, D. Logothetis, G. Siganos, J. Lería, J. Lorenzo, M. Ripeanu, K. Beznosov, and H. Halawa, "Ìntegro: Leveraging victim prediction for robust fake account detection in large scale OSNs," *Comput. Secur.*, vol. 61, pp. 142–168, Aug. 2016.

[14] S.-C. Shi, P. Chen, P.-H. Yuan, C. Hou, and H.-X. Ming, "The prediction of offender identity using decision-making tree algorithm," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, vol. 2, Jul. 2018, pp. 405–409.

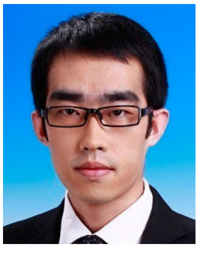
[15] M. Lim, A. Abdullah, N. Jhanjhi, and M. K. Khan, "Situation-aware deep reinforcement learning link prediction model for evolving criminal networks," *IEEE Access*, vol. 8, pp. 16550–16559, 2020.

[16] R. Kumar and B. Nagpal, "Analysis and prediction of crime patterns using big data," *Int. J. Inf. Technol.*, vol. 11, no. 4, pp. 799–805, Dec. 2019.

[17] P. Das, A. K. Das, J. Nayak, D. Pelusi, and W. Ding, "A graph based clustering approach for relation extraction from crime data," *IEEE Access*, vol. 7, pp. 101269–101282, 2019.
[18] Q. Zhang, P. Yuan, Q. Zhou, and Z. Yang, "Mixed spatial-temporal characteristics based crime hot spots prediction," in *Proc. IEEE 20th Int. Conf. Comput. Supported Cooperat. Work Design (CSCWD)*, May 2016, pp. 97–101.
[19] Y. Mei and F. Li, "Predictability comparison of three kinds of robbery crime events using LSTM," in *Proc. 2nd Int. Conf. Data Storage Data Eng. (DSDE)*, 2019, pp. 22–26.
[20] B. J. Jefferson, "Predictable policing: Predictive crime mapping and geographies of policing and race," *Ann. Amer. Assoc. Geographers*, vol. 108, no. 1, pp. 1–16, Jan. 2018.
[21] B. Wang, P. Yin, A. L. Bertozzi, P. J. Brantingham, S. J. Osher, and J. Xin, "Deep learning for real-time crime forecasting and its ternarization," *Chin. Ann. Math. B*, vol. 40, no. 6, pp. 949–966, Nov. 2019.
[22] A. Ghazvini, S. N. H. S. Abdullah, M. K. Hasan, and D. Z. A. B. Kasim, "Crime spatiotemporal prediction with fused objective function in time delay neural network," *IEEE Access*, vol. 8, pp. 115167–115183, 2020.
[23] S. D. Johnson, L. Summers, and K. Pease, "Offender as forager? A direct test of the boost account of victimization," *J. Quant. Criminol.*, vol. 25, no. 2, pp. 181–200, Jun. 2009.
[24] L. E. Cohen and M. Felson, "Social change and crime rate trends: A routine activity approach," *Amer. Sociol. Rev.*, vol. 44, no. 4, pp. 588–608, 1979.
[25] R. V. G. Clarke and M. Felson, *Routine Activity and Rational Choice*. Piscataway, NJ, USA: Transaction Publishers, 1993.
[26] N. B. Davies, J. R. Krebs, and S. A. West, *An Introduction to Behavioural Ecology*. Hoboken, NJ, USA: Wiley, 2012.
[27] D. K. Rossmo, *Geographic Profiling*. Boca Raton, FL, USA: CRC Press, 1999.
[28] S. D. Johnson and K. J. Bowers, "The burglary as clue to the future: The beginnings of prospective hot-spotting," *Eur. J. Criminol.*, vol. 1, no. 2, pp. 237–255, Apr. 2004.
[29] C. P. Haberman and J. H. Ratcliffe, "The predictive policing challenges of near repeat armed street robberies," *Policing A, J. Policy Pract.*, vol. 6, no. 2, pp. 151–166, Jun. 2012.
[30] S. Block and S. Fujita, "Patterns of near repeat temporary and permanent motor vehicle thefts," *Crime Prevention Community Saf.*, vol. 15, no. 2, pp. 151–167, May 2013.
[31] W. Wells, L. Wu, and X. Ye, "Patterns of near-repeat gun assaults in houston," *J. Res. Crime Delinquency*, vol. 49, no. 2, pp. 186–212, May 2012.
[32] J. Sturup, M. Gerell, and A. Rostami, "Explosive violence: A near-repeat study of hand grenade detonations and shootings in urban Sweden," *Eur. J. Criminol.*, vol. 25, no. 4, 2019, Art. no. 1477370818820656.
[33] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 1655–1661.
[34] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*. [Online]. Available: http://arxiv.org/abs/1609.02907
[35] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 7444–7452.
[36] Y. Kong, L. Li, K. Zhang, Q. Ni, and J. Han, "Attention module-based spatial-temporal graph convolutional networks for skeleton-based action recognition," *J. Electron. Imag.*, vol. 28, no. 4, 2019, Art. no. 043032.
[37] X. Geng, Y. Li, L. Wang, L. Zhang, Q. Yang, J. Ye, and Y. Liu, "Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, Jul. 2019, pp. 3656–3663.
[38] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 630–645.
[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

XINGE HAN is currently pursuing the master's degree with the People's Public Security University of China. She has participated in several national academic projects and published one peer-reviewed articles. Her research interest includes deep learning crime risk analysis.

XIAOFENG HU received the Ph.D. degree in nuclear science and technology from Tsinghua University, in 2014. He is currently an Associate Professor with the People's Public Security University of China. His research interests include risk assessment techniques, public safety and emergency management, machine learning, and big data analysis.



HUANGGANG WU received the M.A. degree from Beijing Language and Culture University. He is currently working as a Lecturer with the People's Public Security University of China. His current research interests include international policing and police culture.

BING SHEN is currently pursuing the master's degree with the People's Public Security University of China. He has participated in four national academic projects and published one peer-reviewed articles. His research interest includes machine learning crime risk analysis.

JIANSONG WU received the Ph.D. degree from the Institute for Public Safety Research, Tsinghua University, China. From 2011 to 2012, he was a Visiting Researcher with the Department of Civil Engineering, Johns Hopkins University, USA. He is currently working as a Full Professor with the School of Emergency Management and Safety Engineering, China University of Mining and Technology, Beijing. His main research interests include risk assessment, urban safety and emergency management, occupational safety and health, and CFD modeling.