Deep Multi-view Spatio-Temporal Network for Urban Crime Prediction

Usama Salama[1⋆] Xiaocong Chen[1⋆], Lina Yao[1], Hye-Young Paik[1], Xianzhi Wang[2]

[1] University of New South Wales, Sydney, NSW 2052, Australia
{u.salama, xiaocong.chen, lina.yao, h.paik}@unsw.edu.au
[2] University of Technology Sydney, Ultimo, NSW 2007, Australia
xianzhi.wang@uts.edu.au

Abstract. Crimes sabotage various societal aspects, such as social stability, public safety, economic development, and individuals' quality of life. To accurately predict crime occurrences can not only bring the peace of mind to individuals but also help distribute and manage police resources effectively by authorities. We aim to take into account plenty of environmental factors, such as data collected from Internet of Things (IoT) devices and social networks to predict crimes at city or a finer level. To this end, we propose a deep-learning-based spatio-temporal multi-view model, which explores the relationship between tweets, weather (a type of sensory data) and crime rate, for effective crime prediction. Our extensive experiments on a four-month crime dataset (covering 77 communities, 22 crime types, and 120 days) of Chicago city show that our model can achieve improvement over 19 out of 22 crime types (up to 6.7% in homicide). We also collect the corresponding weather information for different regions of Chicago city to support the crime prediction. Our experiments demonstrate that weather information can improve the performance of the proposed method.

Keywords: Spatial-Temporal Learning, Deep Neural Networks, Crime Prediction

1 Introduction

Crime rates are reportedly rising continuously in several countries and regions. High crime rates undermine a society in multiple aspects, such as public safety, urban stability, medical expenditure, and economic development. It has, therefore, become a key issue to predict crime hotspots (e.g., the places and days when the crime rate would likely to go up) as a key effort on preventing crime effectively. To predict crime occurrences accurately will not only provide timely information for the general public but also assist in allocating police resources more efficiently.

⋆ equal contribution

Several research efforts have contributed to crime prediction. Wang et al. [15] propose a negative binomial regression model to infer a certain community's crime rate based on the characteristics of this community as well as the crime rates in other communities. They introduce two new types of data (Points of Interest and Taxi flows) and explore the correlation the relationship between these two features and crime rate. However, this work estimates crime rate using statistical crime records from other communities in the same year and thus cannot forecast crimes. Gerber et al. [7] combine twitter features and tradition kernel density estimate (KDE) for crime prediction. They employ logistic regression to utilize these features for crime probability estimate and apply distance-weighted spatial interpolation to smooth boundary. Their experiments confirm twitter information can improve the prediction performance in most crime types (19 out of 22).

In light that various features aside from topics (e.g., sentiment index) can be extracted from twitter and can play an important role in crime prediction, Al et al. [1] improve KDE by utilizing of spatio-temporally tagged Twitter posts for inferring micro-level movement patterns. They add temporal information and routine activity patterns inferred from social media as an supplementary feature for crime prediction. Zhao et al. [18] propose a framework TCP to capture temporal-spatial correlation in urban data for crime prediction. They collect a series of statistical information (e.g., historical crime records, check-ins, point-of-interests) as features and further model intra-region temporal correlation and inter-region spatial correlation as regularization for the loss function. Meanwhile, previous research has shown weather influences violent behavior, e.g., a higher temperature generally leads to a higher crime rate [11]. Although previous studies have tried a series of methods to improve crime prediction performance, there is still a large space for improvement:

- Twitter has shown a great potential for event prediction [12]. However, instead of simply using topical and temporal features, a set of context-specific high quality features should be exploited from tweets.
- Deep learning models have demonstrated capability of learning the complex relationships from multi-source data in a variety of applications. It can be adapted for exploiting latent relationships and patterns from different sources of data.
- Crime occurrences might be sparse for certain crime types, resulting in an imbalanced dataset. Hence, a methodology is required to alleviate this problem to sustain an accurate prediction.
- Taxi-flow and geographical distance between different regions can provide a useful spatial correlation to help predict crime, which, however, it yet to be utilized in an efficient manner.
- How the magnitude and spatial distribution of criminal activity is affected by climatic conditions remains largely unexplored in the literature.

Targeting at the above points, we propose a spatial-temporal multi-view model for crime prediction, which utilizes deep learning models to exploit latent correlations and patterns from social and statistical information while simultaneously taking both temporal and spatial relationship into consideration. In summuary, we make the following main contributions in this paper:

– We present a unified model that simultaneously exploit latent relationships from the social information (tweets), statistical information (crime records), and weather information (rain or not) for crime prediction. In particular, we extract a set of high-quality tweet-specific features (emotional, criminal-related, and topical features) for the prediction.
– We efficiently utilize spatial relationships (taxi-flow and geographical distance) between different regions in our model by transforming them into regularization in the loss function. Besides, we utilizes GANs to generate real-time weather information to help predict crimes.
– We propose an iterative Smote-Tomeklinks method, which effectively mitigate the class imbalance problem caused from sparse occurrences in several crime types. Our extensive experiments in real dataset shows that our model outperforms a series of baseline and state-of-the-art methods.

2 Problem Formulation

We aim to predict crime occurrences in Chicago. To this end, we first partition the entire city into different regions. There are two common region-partition methods in previous studies: some [1, 7, 18] manually divide the entire city into disjoint regions with various grid sizes (e.g., 2km $\times$ 2km); others [15] use well-defined, historically recognized community areas. In our work, we divided the city of Chicago to eight regions based on eight main climate stations for Chicago city which are located in Aurora Municipal airport, Botancical garden, Midway airport, Northerly island, Ohare international airport, Palwaukee airport, Waukegan regional airport, and West Chicago Dupage airport. we used Latitude and longitude to scope region partition to guarantee weather data for the region is the closest to the data collected from each climate station. After region partition, we used latitude and longitude to divide crime data to similar regions. We then matched the crime location using latitude and longitude to the latitude and longitude range of each weather station region.

We analyze the statistical crime data of Chicago and find that a certain crime type generally occurs 1 or 0 times in a certain community on a specific day, with a very low possibility of appearing multiple times. In this context, we transform crime prediction problem into a binary-classification task (happen or not happen). Specifically, we denote by $X_{i_k} \in R^{T_1}$ the historical observed records of a crime type k with time window T_1 , $T_i \in R^{T_2}$ the features extracted from tweets posted in community i during time window T_2, W_k^G the geographic distances between community i and the other 76 communities, and finally, W_k^T the taxi flow from the other 76 communities.

3 Data Collection and Feature Extraction

In the past few years, the American government has tried to record crimes precisely to help law enforcement analyze crime patterns and forecast crime hotspots in various cities. Among all the cities in the US, Chicago is ranked third in population (2.7 million) with a high crime rate in serious offenses, such as murders, robberies, aggravated assaults, and property crimes. The 'Chicago Data Portal'[3] provides a complete listing of data categories about this city including taxi trips, parks and recreations, and education, including detailed and immediate-update criminal records. Similar to [7], we collected information on all crimes ($n = 374,043$) documented for the whole year of 2011. Each crime record contains important information, suhc as criminal type (1 out of 30 types, eg. theft, battery and assult), date, description, location (latitude, logitude), and location description. After data cleaning and preparation, we calculate crimes corresponding occurrence frequencies, as shown in Table 1. According to our statistical results, the occurrence frequencies of different crime types fluctuate within an wide range (from 4 to 76145).

Table 1: Frequency of crime types in Chicago documented for the year 2011. We exclude the crime types with very low frequencies.

Crime Types	**No.**	**Crime Types**	**No.**
THEFT	76145	BATTERY	72841
CRIMINAL DAMAGE	43341	DECEPTIVE PRACTICE	11331
ASSAULT	24326	OTHER OFFENSE	21844
NARCOTICS	37264	BURGLARY	20540
VEHICLE THEFT	21278	ROBBERY	14339
CRIMINAL TRESPASS	9722	WEAPONS VIOLATION	3248
CHILDREN OFFENSE	1671	SEXUAL ASSAULT	1319
PUBLIC PEACE VIOLATION	2127	PUBLIC OFFICER INTERFERENCE	320
SEX OFFENSE	1667	PROSTITUTION	4595
HOMICIDE	3751	ARSON	786
LIQUOR LAW VIOLATION	1284	KIDNAPPING	727
GAMBLING	810	INTIMIDATION	221

Features From Social Media We extract three types of features from tweets posted for each training window: emotional features, crime-related features, and topic features. Previous study [7] has shown the importance of topic features from tweets for crime prediction. Besides, we believe other features mined from the tweet content may also play an important role, e.g., the emotional and sentiment index from tweets in a certain community can provide strong supplementary information for crime prediction. We describe each feature type, respectively, as follows: 1) We analyze Emotional Features (F_{E_M}) in

[3] https://data.cityofchicago.org

terms of emotional states, emotional intensity, sentiment score and content polarity. We measure emotional states using 10 emotional terms in tweets, such as 'hate', 'joy', 'sadness', 'anger', and 'nervousness'. Each term is estimated a score using the Empath API [6]. We measure Polarity as a score within the range of (0, 1) by analyzing the whole tweet using the Textblob API. Additionally, we also measure the positive and negative index (ranging from 0 to 1). 2) Regarding crime-related features, we take crime-related indexes from tweets into consideration. We choose up to 12 indexes, such as 'violence', 'aggression', 'dispute', 'Swearing Terms', 'fight', and 'weapon'. Similar to emotional features, these crime indexes also range from 0 to 1 with higher values indicating stronger relationships. 3) Regarding Topical Features F_T, we assume the topic distribution is consistent within each community and extract the topic distribution from a pseudo-document formed by all the filtered tweets from each community during a time window. We firstly exclude non-English tweets and tokenize each tweet using TweetTokenizer [4]. Then, we remove stop words and low frequency words. Differing from previous work [7], we use the original latent Dirichlet Allocation (LDA) to extract topics over unigrams and biterm topic model (BTM) [16] to extract topics over biterms, respectively. We set the number of topics for both topic model to 15. After combining the unigram and biterm topic features together, we get a 30 dimensinal topical feature vector. By assuming that the unigram topical feature and bigram topical feature can be generated from the same hidden representation, we apply a robust auto-encoder [18] to compress the joint topical feature into 15 dimensions. The process is shown in Figure 2 and details of robust auto-encoder can be found in zhou's work. The effectiveness of the extracted latent feature is shown by our experiments in Section 4.1. Finally, we get a 40-dimensional feature vector after feature extraction and partial feature transformation.

Features from Previous Criminal Record Spatial Correlations between Communities Previous studies have shown that the crime rate at one location is highly correlated with nearby locations [9, 15]. Therefore, we take geographical relationships into consideration in crime prediction. Different from (Wang et al. 2016), we form a geo-matrix $W^G(77 \times 77)$ by setting W^G_{ii}=0 and W^G_{ij} as the euclidean distance between the centroids of communities i and j (W^G_{ij} is symmetric) [5]. Then, all the elements W^G_{ij} of the matrix are normalized over the row W^G_i. Besides geographical influence, taxi flows can also build a bridge between different communities and influence the crime rate [10, 15]. Taking this into account, we build a taxi-flow matrix W^T in a similar manner as W^G, where each W^T_{ii} is set as zero and W^T_{ij} (W^T_{ij} is asymmetric) is calculated by summing up all the taxi flows from community j to i. **Over-sampling and Data Cleaning** Before feeding the extracted features into our model, we conduct data interpolation in the minority class of data to solve the imbalance problem. Previous study [2] has conducted rich analyses of existing over-sampling methods and propose a new variant Smote+Tomek links, which shows excellent

[4] https://www.nltk.org/api/nltk.tokenize.html
[5] https://github.com/brandonxiang/geojson-python-utils

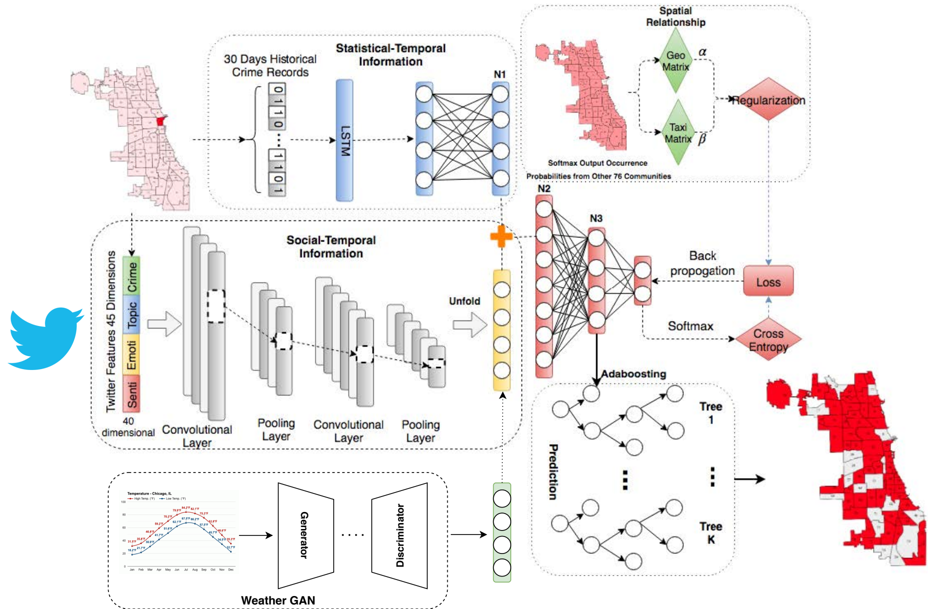


Fig. 1: The structure of the proposed model which utilize three different types of data: crime record, twitter and weather data.

performance in their experiments. The original Smote-Tomeklinks successfully combine the advantages of Smote (over-sampling) [3] and Tomek-links [14] (data-cleaning)—the method first oversamples minority data samples with Smote and then identifies and removes Tomek links. Details of Smote and Tomeklinks can be found in [3, 14]. However, this method still has an disadvantage if a serous imbalance exists in a dataset because the noises in the dataset can actually influence the final calssification performance. More specifically, Smote might create a large amount of artificially invented samples, which contain considerable noises that cannot be removed by Tomeklinks. In this regard, we proposed an iterative Smote-Tomeklinks model, which employ Smote and Tomeklinks iteratively to clean out the noises in a timely manner. The overall process is shown in Figure 2. The key difference of our method from the original Smote+Tomeklinks method is that we conduct Smote and Tomeklinks iteratively for T times (until reaching the criteria). Besides, to avoid excessive cleaning to the data in initial stages, we changed the rule of Tomeklinks in the first $T-1$ times by removing the identified pairs under a certain probability (e.g., 0.35 in our work).

4 Methodology

We propose a spatial-temporal multi-view model for predicting crime occurrences, the overall process of which is shown in Figure 1. The model consists of three main parts: Weather information generation, Statistical features, and Spatial relationship regularization.

4.1 Information Gathering

Online-information When extracting different kinds of features from tweets during the past time window (i.e., one week in this paper), we aim to exploit the latent relationships between theses features for crime prediction. Previous study [5,13] have shown the appealing ability of convolutional neural network in classification tasks. Therefore, we build a one-dimensional convolutional neural network for online information utilization. Our 1DCNN model consists of four layers: two convolutional layers followed by two pooling layers. The twitter features will be shared for all the crime types in a certain community.
Offline-information We build a Long short term memory based network (LSTM), which is suitable for processing time-serial data, to fully leverage historical crime observation. We feed a time window (which is set as one month) of historical crime data (a certain crime type e.g., theft) in a target community into an LSTM network.

The online information and offline information are exploited via 1DCNN and LSTM, respectively, and then unfolded into two one-dimensional vectors to be combined and fed into three fully connected layers, together with weather information for the final classification.

Weather Information We additionally integrate weather information into our model to improve the performance. Specifically, we employ GANs [8] to generate the corresponding weather information. GANs can be formulated as a distribution-matching problem, which aims to minimize the distance between the learned weather distribution and an expected distribution. TO relive GANs from the difficutly of handle time-series weather data we adopt TimeGAN [17]. The loss function of the TimeGAN $\mathcal{L}_G$ has two components $\mathcal{L}_S, \mathcal{L}_U$ as specified below:

$$\mathcal{L}_S = \mathbb{E}_{s,x_1:T\sim p}[\sum_t \|h_t - g_{\mathcal{X}}(h_S, h_{t-1}, z_t)\|_2]$$

$$\mathcal{L}_U = \mathbb{E}_{s,x_1:T\sim p}[\log y_S + \sum_t \log y_t] + \mathbb{E}_{s,x_1:T\sim \hat{p}}[\log(1-\hat{y}_S) + \sum_t \log(1-\hat{y}_t)]$$

where s, x_1 is the input following the distribution $p(S, X_{1:T})$, which is time-series data. $g_{\mathcal{X}}(h_S, h_{t-1}, z_t)$ is the output state of a LSTM network. y_S is the classification result from a softmax function. Hence, we define the final loss function for GAN as:

$$\mathcal{L}_G = \min_{\theta_g}(\eta\mathcal{L}_S + \max_{\theta_d}\mathcal{L}_U) \tag{1}$$

For temporal information, we use the binary cross-entropy as the loss function, as shown below:

$$\mathcal{L}_T = -\sum_{i=0} y_i \log(y'_i) + \alpha|y_1 - W_G^K Y'_1| + \beta|y'_1 - W_k^T Y'_1| \tag{2}$$

where W_i^G and W_i^T are the i-th row of geo-matrix and taxi flow matrix defined in the above section, Y_1' is a 77-dimensional vector in which each element Y_{1j}' denotes the predictive occurrence probability chosen from the softmax output of community j (Y_{1i}' is set to 0), α and β are the tuning weights for geographic and taxi flow influence, respectively. In particular, the second and third items can be regraded as a regularization for loss function to increase the robustness and to prevent overfitting; α and β will be tuned as hyper-parameters.

5 Experiments

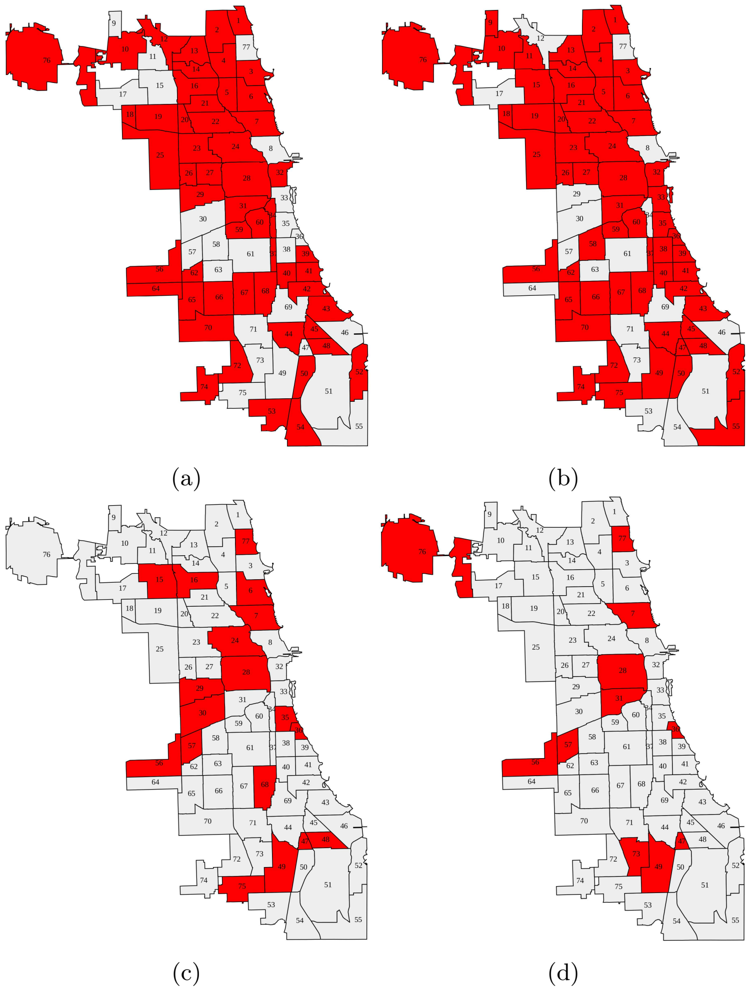
We collect $152,460(90 \times 77 \times 22)$ samples of crime records spanning four months in Chicago. We set the time window to one month for historical crime data and one week for weather data. We train our model for each crime separately, as each crime has its own unique pattern and it does not improve the performance to consider training samples from other crime types. We split the dataset 90×77) of each crime type into training samples (5,005) and testing samples (1,935) by time. Therefore, testing samples contain the more recent data that training samples. Furthermore, we also conduct sampling and data cleaning for crime types, regarding which occurrence samples takes less then 20% of the total samples. Specifically, we apply iterative tomek-links to increase the percentage until it exceeds 20%. We grab historical weather information from a public source[6] and transform weather data into binary (indicating rain or not). We finally conduct experiments a machine with 8 GPUs and use accuracy as the evaluation metric.

5.1 Comparison Methods

- KDE(T): This model [7] combines features extracted from historical crime records using kernel density estimation(KDE) with topical features from twitter. Its experimental results shows improvement over 19 out of 26 crime types. We apply their model to our data using the same topical features used by our method to make a fair comparison. As we use natural community as region partition, our method does not require spatial interpolation as this method does.
- KDE(T*): This model [1] is a variant of the KDE(T). It models the temporal feature from twitter in a new way and combine them with historical crime records. We use the recommended settings in the original paper.
- KDE(T*+R): This model is an improved version of KDE(T*) and is also proposed by [1]. This model extracts routine activity patterns from twitter as additional features and feed them with KDE results into a binary classification method.
- TCP: TCP [18] is a framework that captures the intra-region temporal relationship and inter-region spatial relationship for crime prediction. We employ TCP on our features and transform its task from regression into classification. We set two parameters λ and ηH as 1 and 0.4, respectively, for our task.

[6] https://www.wunderground.com/

– LSTM: We directly feed all features to this model for crime prediction, setting the hidden cells of LSTM layer to 70 and three fully-connected layers with 120, 60, 80 neurons following the LSTM layer.
– GRU: Gated Recurrent Unit(GRU) is a a gating mechanism in recurrent neural networks [4]. It is similar to LSTM model but achieves better performance in many tasks. We feed all the features to this model and set hidden states of GRU layer to 70 and two fully-connected layers with 100, 50 neurons.



(a) (b) (c) (d)

Fig. 2: True Occurrences of Theft Over Entire 12th(a); Predictive Occurrences of Theft Over 12th(b); True Occurrences of Robbery Over Entire City on April 12th(c); Predictive Occurrences of Robbery Over Entire City on April 12th(d)

Our comparison results (Table 2) show the proposed method outperform the baseline methods on 19 out of 22 crime types.

Table 2: Experimental results. CD denotes criminal damage; DP denotes deceptive practice; OO denotes other offenses; MVT denotes motor vehicle theft; CT denotes criminal trespass; WV denotes weapons violation; CO denotes children offense; CSA denotes crim sexual assault; PPV denotes public peace violation; PCI denotes public officer interference; LLV denotes liquor law violation; $\neg G$ denotes our model excluding geo-information; $\neg T$ denotes our model excluding taxi-information; $\neg W$ denotes our model excluding weather features.

Crime	**SOTAs**			**DL based**			**Proposed**			
	KDE(T*)	KDE(T*+R)	KDE(T)	TCP	GRU	LSTM	Ours	$\neg G$	$\neg T$	$\neg W$
Theft	0.66	0.69	0.68	0.70	0.68	0.66	**0.74**	0.722	0.729	0.728
Battery	0.72	0.75	0.74	0.73	0.76	0.76	**0.82**	0.777	0.792	0.810
CD	0.66	0.69	0.71	0.68	0.66	0.70	**0.75**	0.720	0.710	0.748
DP	0.68	0.72	0.67	0.71	0.64	0.66	**0.72**	0.677	0.676	0.711
Assault	0.66	0.72	0.73	0.71	0.69	0.70	**0.75**	0.743	0.733	0.741
OO	0.65	0.70	0.71	0.65	0.67	0.71	**0.77**	0.759	0.757	0.766
Narcotics	0.74	0.75	0.75	0.72	0.69	0.72	**0.77**	0.745	0.751	0.760
Burglary	0.67	0.66	0.67	0.71	0.65	0.69	**0.73**	0.701	0.677	0.709
MVT	0.62	0.66	0.67	0.66	0.63	0.62	**0.68**	0.667	0.662	0.681
Robbery	0.71	0.74	0.75	0.73	0.70	0.70	**0.78**	0.770	0.770	0.771
CT	0.66	**0.70**	0.69	**0.70**	0.67	0.67	0.66	0.646	0.621	0.651
WV	0.68	0.70	0.74	0.73	0.72	0.73	**0.77**	0.766	0.745	0.763
CO	0.59	0.62	0.60	0.63	0.60	0.61	**0.65**	0.641	0.631	0.641
CSA	0.66	0.67	0.65	0.64	0.62	0.63	**0.71**	0.684	0.681	0.702
PPV	0.60	0.63	0.65	0.64	0.61	0.62	**0.66**	0.655	0.658	0.643
PCI	0.69	0.72	**0.76**	0.74	0.70	0.71	**0.76**	0.732	0.743	0.752
SO	0.54	0.58	0.56	0.57	0.54	0.54	**0.59**	0.560	0.550	0.581
Prostitution	0.74	0.76	0.81	0.79	0.80	0.81	**0.87**	0.841	0.851	0.862
Homicide	0.61	0.59	0.62	0.63	0.61	0.63	**0.70**	0.674	0.662	0.692
Arson	0.53	0.57	0.57	0.54	0.51	0.51	**0.58**	0.544	0.533	0.571
LLV	0.63	0.62	**0.64**	0.59	0.56	0.56	0.63	0.581	0.575	0.620
Kidnapping	0.47	0.51	0.50	0.48	0.49	0.50	**0.53**	0.520	0.511	0.521

5.2 Hyper-parameters Tuning

We set the hyperparameters of our model as follows: the number of hidden cells in LSTM layer as 30, filter number of convolutional layer as 10, filter length of filter of convolutional layer, number ofneurons N1 as 60, number of neurons N2 as 80, number of neurons as 120, number of estimators in adaboosting as 5, learning rate of adaboosting as 0.7. For the GAN, we fix both the temporal correlation and feature correlations at 0.2.

5.3 Ablation Study

We conduct ablation study to investigate the effect of weather data, taxi information, and geo-information on our model's performance. The results are shown in the fourth column in Table 2.

6 Conclusion and Future work

Crime prediction finds its root and rationale in the fact that many criminals tend to commit the same types of crimes proven successfully in similar time and locations. This paper presents a preliminary investigation of social and weather factors for crime activity prediction, and our experiments demonstrate the model's ability to improve the forecast of crime occurrences, along with a clear correlation between crime and weather forecasting. There are plenty of ways to improve the efficiency and accuracy of crime prediction using IoT sensors. In this regard, our future work will include obtaining hourly weather data instead of daily data from IoT sensors in small regions, using crime data that include time and postcode to match with sensor data, and adding non-weather sensor data (e.g., number of people passing, traffic at certain times, and noise level) to make more accurate predictions.

References

1. Al Boni, M., Gerber, M.S.: Predicting crime with routine activity patterns inferred from social media. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 001233–001238. IEEE (2016)
2. Batista, G.E., Prati, R.C., Monard, M.C.: A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD explorations newsletter **6**(1), 20–29 (2004)
3. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research **16**, 321–357 (2002)
4. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
5. Ding, X., Liu, T., Duan, J., Nie, J.Y.: Mining user consumption intention from social media using domain adaptive convolutional neural network. In: AAAI. vol. 15, pp. 2389–2395 (2015)
6. Fast, E., Chen, B., Bernstein, M.S.: Empath: Understanding topic signals in large-scale text. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. pp. 4647–4657 (2016)
7. Gerber, M.S.: Predicting crime using twitter and kernel density estimation. Decision Support Systems **61**, 115–125 (2014)
8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)

9. Gorman, D.M., Speer, P.W., Gruenewald, P.J., Labouvie, E.W.: Spatial dynamics of alcohol availability, neighborhood structure and violent crime. Journal of studies on alcohol **62**(5), 628–636 (2001)
10. Kadar, C., Pletikosa, I.: Mining large-scale human mobility data for long-term crime prediction. EPJ Data Science **7**(1), 26 (2018)
11. Kenrick, D.T., MacFarlane, S.W.: Ambient temperature and horn honking: A field study of the heat/aggression relationship. Environment and behavior **18**(2), 179–191 (1986)
12. Ning, X., Yao, L., Benatallah, B., Zhang, Y., Sheng, Q.Z., Kanhere, S.S.: Source-aware crisis-relevant tweet identification and key information summarization. ACM Transactions on Internet Technology (TOIT) **19**(3), 1–20 (2019)
13. Ning, X., Yao, L., Wang, X., Benatallah, B.: Calling for response: Automatically distinguishing situation-aware tweets during crises. In: International Conference on Advanced Data Mining and Applications. pp. 195–208. Springer (2017)
14. Tomek, I., et al.: An experiment with the edited nearest-nieghbor rule. (1976)
15. Wang, H., Kifer, D., Graif, C., Li, Z.: Crime rate inference with big data. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. pp. 635–644 (2016)
16. Yan, X., Guo, J., Lan, Y., Cheng, X.: A biterm topic model for short texts. In: Proceedings of the 22nd international conference on World Wide Web. pp. 1445–1456 (2013)
17. Yoon, J., Jarrett, D., van der Schaar, M.: Time-series generative adversarial networks. In: Advances in Neural Information Processing Systems. pp. 5508–5518 (2019)
18. Zhao, X., Tang, J.: Modeling temporal-spatial correlations for crime prediction. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. pp. 497–506 (2017)