Spatial-Temporal Augmentation for Crime Prediction (Student Abstract)

Hongzhu Fu[1], Fan Zhou[1,3], Qing Guo[4], Qiang Gao[2,3,*]

[1]University of Electronic Science and Technology of China
[2]Southwestern University of Finance and Economics
[3]Kash Institute of Electronics and Information Industry
[4]School of Computer Science and Technology, Beijing Institute of Technology
fuhongzhu@std.uestc.edu.cn,fan.zhou@uestc.edu.cn, 3120195496@bit.edu.cn, qianggao@swufe.edu.cn

Abstract

Crime prediction stands as a pivotal concern within the realm of urban management due to its potential threats to public safety. While prior research has predominantly focused on unraveling the intricate dependencies among urban regions and temporal dynamics, the challenges posed by the scarcity and uncertainty of historical crime data have not been thoroughly investigated. This study introduces an innovative spatial-temporal augmented learning framework for crime prediction, namely STAug. In STAug, we devise a CrimeMix to improve the ability of generalization. Furthermore, we harness a spatial-temporal aggregation to capture and incorporate multiple correlations covering the temporal, spatial, and crime-type aspects. Experiments on two real-world datasets underscore the superiority of STAug over several baselines.

Introduction

Accurately predicting criminal activities has emerged as a critical endeavor within the realm of urban safety and management. The presence of various criminal incidents, such as robberies and burglaries, poses a continuous threat to the well-being of both individuals and society as a whole. Recent studies (Xia et al. 2021; Zhao et al. 2022; Gao et al. 2023) mainly focused on exploring the multiple crime dependencies with sparse crime data. However, the uncertain interactive semantics are under-explored. To address the above concern, as shown in Fig. 1, we present a novel solution called STAug for crime prediction. STAug first operates a CrimeMix module to effectively enrich samples from vicinity distribution and increase adversarial robustness. Then, it operates the spatial-temporal aggregation that primarily investigates the temporal dynamics, crime type relations, and spatial correlations underlying crime cases.

Solution: STAug

Problem Definition. Given historical crime data $\mathbf{X} \in \mathbb{R}^{R\times T\times C}$, with R, T, and C signifying the number of grid-scale regions, time slots, and crime types, respectively. Our objective is to develop a model that can estimate the future situation of crime incidents for various crime types in each region for the upcoming time slot $T+1$, which can be denoted as $\hat{\mathbf{X}}_{T+1} \in \mathbb{R}^{R\times C}$.

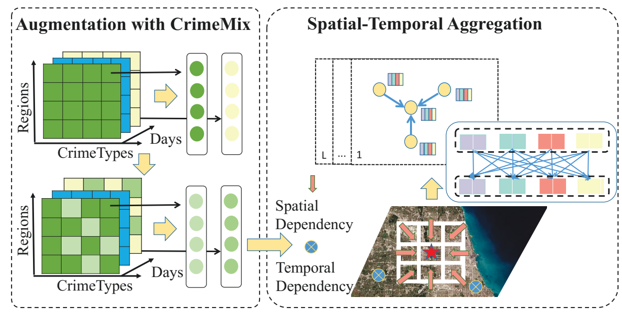

Figure 1: The architecture of STAug.

Augmentation with CrimeMix. Based on prior relevant research, it is evident that crime data exhibits a high degree of sparsity in its distribution, which significantly impacts the training of models. Consequently, we propose a novel module named CrimeMix to enhance the crime model's performance, which draws inspiration from Mixup (Zhang et al. 2018). The primary purpose is to enhance the robustness and generalization capabilities of models by generating new training examples through a combination of existing crime data time, which can be denoted as follows:

$$\tilde{x}_{t,c} = \lambda x_{t_1,c} + (1-\lambda)x_{t_2,c}, \tag{1}$$

$$\tilde{x}_{r,t} = \lambda x_{r,t_1} + (1-\lambda)x_{r,t_2}, \tag{2}$$

where $x_{t_1,c}$,$x_{t_2,c}$ are raw urban crime vectors, and x_{r,t_1},x_{r,t_2} are crime multi-label vectors. $\lambda \in [0,1]$ is distributed according to a Beta distribution: $\lambda \sim \beta(\alpha,\alpha)$. In this way, we can produce rich samples to boost crime pattern learning.

Crime Embedding. We first prepare dense representations for each crime type. We start by generating random initial embeddings $\mathbf{e}_c \in \mathbb{R}^d$ for each crime type c. These initial embeddings serve as a foundation. Then, we create specific representations $\mathbf{e}_{r,t,c} \in \mathbb{R}^d$ for each instance, capturing the crime pattern associated with the c-th type in region r during time slot t. These representations are computed as follows:

$$\mathbf{e}_{r,t,c} = \frac{\tilde{\mathbf{X}}_{r,t,c} - \mu}{\sigma} \cdot \mathbf{e}_c, \tag{3}$$

*Corresponding author (qianggao@swufe.edu.cn).

where μ and σ refer to the mean and standard deviation of the entire tensor $\tilde{\mathbf{X}}$ for normalization. The resulting four-dimensional tensor $\mathbf{H} \in \mathbb{R}^{R\times T\times C\times d}$ serves as input for the next step of our analysis.

Spatial-Temporal Aggregation. To capture correlations across different regions and crime types, we apply routing and attention mechanisms. Specifically, for each time slot t, the general pipeline can be denoted as follows:

$$\mathrm{Attn}\left(\mathbf{H}_{i,t,c}, \mathbf{H}_{j,t}\right) = \|_{m=1}^{M} \sum_{c'=1}^{C} \gamma_{c,c'}^{m} \cdot \mathbf{V}^{m} \mathbf{H}_{j,t,c'}, \quad (4)$$

Herein, $\gamma_{c,c'}^{m}$ represents the attention score between two crime types, capturing the dependency degree between crime type c in region r_i and crime type c' in region r_j. $\mathbf{V}^{m} \in \mathbb{R}^{d/M\times d}$ is the trainable parameters regarding $\mathbf{H}_{j,t,c'}$. We use multi-head attention for diverse attention views. Hence, $||$ represents a concatenation operation. Correspondingly, the attention score $\gamma_{c,c'}^{m}$ is computed as:

$$\bar{\gamma}_{c,c'}^{m} = \frac{\left(\mathbf{Q}^{m}\mathbf{H}_{i,t,c}\right)^{\top}\left(\mathbf{K}^{m}\mathbf{H}_{j,t,c'}\right)}{\sqrt{d/M}}, \quad (5)$$

$$\gamma_{c,c'}^{m} = \frac{\exp\left(\bar{\gamma}_{c,c'}^{m}\right)}{\sum_{c'}\exp\left(\bar{\gamma}_{c,c'}^{m}\right)}, \quad (6)$$

where $\mathbf{Q}^{m}$ and $\mathbf{K}^{m} \in \mathbb{R}^{d/M\times d}$ are trainable parameters.

To explore spatial and temporal interactions, we construct a crime graph G, treating each region r as a graph node $v \in V$. We establish edges between neighboring regions to capture spatial correlations and between regions where crimes occur on the same day to incorporate temporal correlation. In detail, an adjacency matrix A representing these correlations can be defined as follows:

$$A_{ij} = \begin{cases} 1, \text{if } v_i \text{ and } v_j \text{ are } \textit{neighboring}; \\ 0, \text{ otherwise.} \end{cases} \quad (7)$$

Herein, the *neighboring* indicates the relation between two nodes is either spatial or temporal relevant. To capture long-distance interactions, we make message propagation as:

$$\mathbf{H}_{i}^{(l+1)} = ReLU\left(\sum_{j=1}^{R} A'_{i,j} \cdot \mathrm{Attn}(\mathbf{H}_{i}^{(l)}, \mathbf{H}_{j}^{(l)})\right), \quad (8)$$

where $\mathbf{H}_{i}^{(l+1)}$ is the representation after $l \in L$ iterations. Besides, $A' = D^{-1/2}AD^{1/2}$ is a normalized adjacent matrix, where D indicates a diagonal (degree) matrix.

Optimization. Now we sum up the target crime tensor $\mathbf{H}^{L}$ along the time slot dimension to obtain the final representation $\boldsymbol{\kappa} \in \mathbb{R}^{R\times C\times d}$ and produce the next time slot's crime result $\hat{\mathbf{X}}_{T+1}$ with the Sigmoid function. The crime prediction task minimizes the objective as follows:

$$\mathcal{L} = -\sum_{t}^{T} \delta\left(\mathbf{X}_t\right)\log\hat{\mathbf{X}}_t + \bar{\delta}\left(\mathbf{X}_t\right)\log\left(1-\hat{\mathbf{X}}_t\right) + \eta\|\mathbf{\Theta}\|_2^2, \quad (9)$$

where $\delta\left(\cdot\right)$ and $\bar{\delta}\left(\cdot\right)$ are element-wise positive and negative indicator functions, respectively (Xia et al. 2021). The last term is L_2 regularization and η denotes the decay weight.

Data	**NYC-Crimes**		**Chicago-Crimes**	
Crime Types	Burglary	Robbery	Theft	Battery
Number	31,799	33,453	124,630	99,389
Crime Types	Assault	Larceny	Damage	Assault
Number	40,429	85,899	59,886	37,972

Table 1: The description of urban crime datasets.

Method	**NYC**		**CHI**	
	micro	macro	micro	macro
DeepCrime	0.5727	0.5713	0.5186	0.5174
ST-SHN	0.6154	0.6144	0.6390	0.6410
ST-DPL	0.6465	0.6413	0.7237	0.7296
STAug	**0.6573**	**0.6567**	**0.7361**	**0.7356**

Table 2: Overall performance on two datasets.

Experiment

We evaluate our proposed method using the crime data collected from Chicago (CHI) and New York City (NYC). Each record in the dataset includes information about crime category, timestamp, and geographical coordinates. Table 1 records the statistics of them. We select DeepCrime (Huang et al. 2018), ST-SHN (Xia et al. 2021) and ST-DPL (Gao et al. 2023) as our baselines. We adopt two widely used metrics for performance evaluation, including micro-F_1 and macro-F_1 (Huang et al. 2018; Xia et al. 2021).

Performance Comparison. Table 2 presents the main results. Note that we use *micro* and *macro* in Table 2 to represent micro-F_1 and macro-F_1, respectively. Overall, we can observe that our model achieves the best gains in performance across all datasets. Compared with baselines, our approach not only proposes a novel spatial data argumentation to alleviate the sparsity issue but also considers the complex spatial-temporal dependencies between different types of crime patterns.

Acknowledgments

This work was supported by the Natural Science Foundation of Sichuan Province under Grant No.2023NSFSC1411.

References

Gao, Q.; Fu, H.; Wei, Y.; Huang, L.; Liu, X.; and Liu, G. 2023. Spatial-Temporal Diffusion Probabilistic Learning For Crime Prediction. In *KSEM*, 263–275.

Huang, C.; Zhang, J.; Zheng, Y.; and Chawla, N. V. 2018. DeepCrime: Attentive hierarchical recurrent networks for crime prediction. In *CIKM*, 1423–1432.

Xia, L.; Huang, C.; Xu, Y.; Dai, P.; Bo, L.; Zhang, X.; and Chen, T. 2021. Spatial-Temporal Sequential Hypergraph Network for Crime Prediction with Dynamic Multiplex Relation Learning. In *IJCAI*, 1631–1637.

Zhang, H.; Cisse, M.; Dauphin, Y. N.; and Lopez-Paz, D. 2018. mixup: Beyond Empirical Risk Minimization. In *ICLR*, 1–13.

Zhao, X.; Fan, W.; Liu, H.; and Tang, J. 2022. Multi-type urban crime prediction. In *AAAI*, 4, 4388–4396.