MiST: A Multiview and Multimodal Spatial-Temporal Learning Framework for Citywide Abnormal Event Forecasting

Chao Huang
University of Notre Dame
chuang7@nd.edu

Chuxu Zhang
University of Notre Dame
czhang11@nd.edu

Jiashu Zhao
JD.com
zhaojiashu1@jd.com

Xian Wu
University of Notre Dame
xwu9@nd.edu

Dawei Yin
JD.com
yindawei@acm.org

Nitesh V. Chawla
University of Notre Dame
nchawla@nd.edu

ABSTRACT

Citywide abnormal events, such as crimes and accidents, may result in loss of lives or properties if not handled efficiently. It is important for a wide spectrum of applications, ranging from public order maintaining, disaster control and people's activity modeling, if abnormal events can be automatically predicted before they occur. However, forecasting different categories of citywide abnormal events is very challenging as it is affected by many complex factors from different views: (i) dynamic intra-region temporal correlation; (ii) complex inter-region spatial correlations; (iii) latent cross-categorical correlations. In this paper, we develop a Multi-View and Multi-Modal Spatial-Temporal learning (MiST) framework to address the above challenges by promoting the collaboration of different views (spatial, temporal and semantic) and map the multi-modal units into the same latent space. Specifically, MiST can preserve the underlying structural information of multi-view abnormal event data and automatically learn the importance of view-specific representations, with the integration of a multi-modal pattern fusion module and a hierarchical recurrent framework. Extensive experiments on three real-world datasets, *i.e.*, crime data and urban anomaly data, demonstrate the superior performance of our MiST method over the state-of-the-art baselines across various settings.

CCS CONCEPTS

• **Information systems → Spatial-temporal systems**.

KEYWORDS

Abnormal Event Forecasting, Deep Neural Networks; Spatial-temporal Data Mining

ACM Reference Format:
Chao Huang, Chuxu Zhang, Jiashu Zhao, Xian Wu, Dawei Yin, and Nitesh V. Chawla. 2019. MiST: A Multiview and Multimodal Spatial-Temporal Learning Framework for Citywide Abnormal Event Forecasting. In *Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3308558.3313730

WWW'19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313730

1 INTRODUCTION

Citywide abnormal events, like crimes (*e.g.*, robbery, felony assault) and urban anomalies (*e.g.*, blocked driveway, noise), may pose tremendous risks to public safety if not timely handled. According to national statistics, urban anomalies cost Americans more than $100 billion direct and indirect loss in each of recent years [39]. Hence, accurate and reliable prediction of abnormal events is in pressing need for data-driven decision makers to alleviate huge life and economic losses. For example, in disaster control, by forecasting future abnormal events, local governments can design better transportation planning and mobility management strategies to prevent severe social riots [34, 48]. Furthermore, in public order maintaining services, understanding latent occurrence patterns of abnormal events at each geographical region in a city is highly important for people's activities modeling and place recommendation tasks [20, 40]. In this paper, we aim to predict abnormal events of different categories at each region of a city ahead of time, thereby leading to significant improvement in societal welfare.

There exist prior studies on detecting geographical anomalies using spatial-temporal data [7, 17, 26]. Most of these studies identify the abnormal events by analyzing the historic traces or movement patterns of the studied objects using statistical and data mining approaches. However, instead of forecasting events in the future, these approaches typically identify them only after they have occurred, which may lead to significant information delay and lack of preparedness to handle the anomalies in an efficient way.

We identify three key challenges of modeling such abnormal event data from multiple views, which motivate the model design:

First, the distributions of abnormal events in the urban space are highly dynamic and vary greatly from one region to anther. In such cases, the abnormal event occurrences are no longer independent across different locations and it is crucial to take the spatial correlations into consideration for the abnormal event forecasting task. In addition, when modeling dynamic pairwise spatial correlations, probabilistic graphical models [23, 24] may fail due to their heavy computational cost with massive parameters based on prior assumed distributions.

Second, the occurrence pattern of abnormal events often involve underlying factors which are dynamically evolving over time. For example, crime causality on weekdays may differ from weekends. Conventional time series forecasting techniques, such as Autoregressive Integrated Moving Average (ARIMA) [29] and Support Vector Regression (SVR) [30], are mostly limited to linear models which depend solely on the single-level periodic patterns. Thus,

these methods can hardly be responsible to temporal dynamics across different time slots with respect to the occurrences of abnormal events.

Third, explicit and implicit influences among different abnormal event categories are ubiquitous in real life. For example, traffic jam at a specific region may be triggered by a robbery occurring in the same region, due to the crowd aggregation and increased patrol in response to the robbery. Therefore, the occurrence of a specific category of abnormal event may stem from not only spatial correlations between regions as well as temporal dependencies across time slots, but also the influences among different categories.

Motivated by the aforementioned challenges, this work proposes a general and flexible framework–Multi-View Deep Spatial-Temporal Networks (MiST)–for learning the predictive structure from the relationships within the multi-view abnormal event data. Specifically, at the first phase, we develop a context-aware recurrent framework to capture temporal dynamics of abnormal event data in different views and provide automatic view-specific representations. At the second phase, to jointly preserve the inter-region correlations and the cross-category influences together with the encoded temporal patterns of multi-dimensional data, we propose a pattern fusion module based on attention mechanisms to promote the collaboration of different views, and automatically capture the contribution of correlated regions, time slots and categories from the corresponding view in the predictive model. To supercharge MiST modeling of summarized occurrence patterns with time-ordered structural information and non-linearities, a conclusive recurrent network module is designed to model the sequential patterns of fused embedding vectors at the final phase. The final conclusive latent representations are fed into a fully connected neural networks to predict the abnormal events in future time slots.

In summary, we highlight our contributions as follows:

- We introduce a new multi-view and multi-modal spatial-temporal learning framework MiST for predicting abnormal events of different categories at each region of a city. MiST maps all the spatial, temporal, and semantic units into a latent space to preserve their cross-modal correlations.
- We develop a multi-modal pattern fusion module that is tailored to cooperate with a hierarchical recurrent framework for learning the latent region-time-category interactions shared in multi-view data, and automatically adjust the correlations from each view in assisting the prediction task.
- Through extensive experiments performed on three real-world abnormal event datasets collected from NYC and Chicago, we show that MiST consistently surpasses several state-of-the-art baselines across various settings.

The rest of this paper is organized as follows. We formally present the problem in Section 2. Section 3 introduces the proposed framework. The experimental results are presented in Section 4. Section 5 summarizes the related works. We finally conclude our work in Section 6.

2 PROBLEM FORMULATION

In this section, we first introduce the preliminaries and the studied problem. Then, we present the overview of our developed framework.

2.1 Preliminaries

Definition 1. **Geographical Region**. *We partition a city with a grid-based map segmentation. In particular, we divide the whole city into $I \times J$ disjointed grids which consists of I rows and J columns with the longitude and latitude information. Each grid is referred to as a geographical region denoted by $r_{i,j}$, where i and j is the index of row and column, respectively. In this work, we use regions as the minimal units to study the abnormal event forecasting problem.*

We define the geographical region set as $R = (r_{1,1}, ..., r_{i,j}, ..., r_{I,J})$, and suppose that there are L abnormal event categories ($C = (c_1, ..., c_l, ..., c_L)$), where C is the category set indexed by l. Given a time window T, we also split T as non-overlapping and consequent time slots ($T = (t_1, ..., t_k, ..., t_K)$), where K denotes the number of time slots which is indexed by k.

Definition 2. **Abnormal Event Data Source**. *Given a region $r_{i,j}$, we use $\mathbf{Y}_{i,j} = (\boldsymbol{y}^1_{i,j}, ..., \boldsymbol{y}^l_{i,j}, ..., \boldsymbol{y}^L_{i,j}) \in \mathbb{R}^{L \times K}$ to denote the observations of all categorical abnormal events during K past time slots at region $r_{i,j}$. In particular, $\boldsymbol{y}^l_{i,j} \in \mathbb{R}^K$ belongs to category c_l and represents the event observed at region $r_{i,j}$ from time slot t_1 to t_K. In $\boldsymbol{y}^l_{i,j}$, each element $y^{l,k}_{i,j}$ is set to 1 if abnormal event of category c_j is observed at region $r_{i,j}$ in time slot t_k and 0 otherwise.*

Problem Statement. Given the historical abnormal event data source of all categories across geographical regions R in a city from time slot t_1 to t_K, the objective of this work is to learn a predictive framework which infers the unknown event occurrence of each category c_l at each region $r_{i,j}$ in h future time slots. Formally, we aim to compute: $(y^{l,(K+h)}_{i,j} | \mathbf{Y}_{i,j} = (\mathbf{y}^1_{i,j}, ..., \mathbf{y}^L_{i,j}));\ i, j \in [1, ..., I], [1, ..., J]$.

2.2 Framework Overview

Our developed MiST is a multi-layer representation learning framework as shown in Figure 1. Before presenting the details, we first introduce the model input and then elaborate the motivations to design our framework MiST from spatial-temporal-categorical views.

Definition 3. **Event Context Tensor**. *Given a target region $r_{i,j}$, we model different categorical abnormal events of its spatially nearby regions in time slot t_k using an event context tensor, $\mathcal{A}^k_{i,j} \in \mathbb{R}^{I \times J \times L}$ with three dimensions denoting I grid rows, J grid columns, and L categories, respectively. Specifically, given the time slot t_k, each element in $\mathcal{A}^k_{i,j,l}$ is set as 1 if there exists abnormal events of category c_l from the corresponding region $r_{i,j}$ in the $I \times J$ grid map. Otherwise, we set the element in $\mathcal{A}^k_{i,j,l}$ as 0.*

Context-aware Recurrent Framework. To characterize intra-region correlations in terms of dynamic properties of abnormal event distributions from temporal view, we develop our context encoder based on Long short-term memory (LSTM) networks to learn latent representations for the elements of the flattened vector from event context tensor $\mathcal{A}$ in each time slot. The representations learned from our LSTM encoder give MiST the ability to model the evolving time-dependent nature of abnormal events by well capturing not only their local temporal contexts but also multi-level periodical patterns.

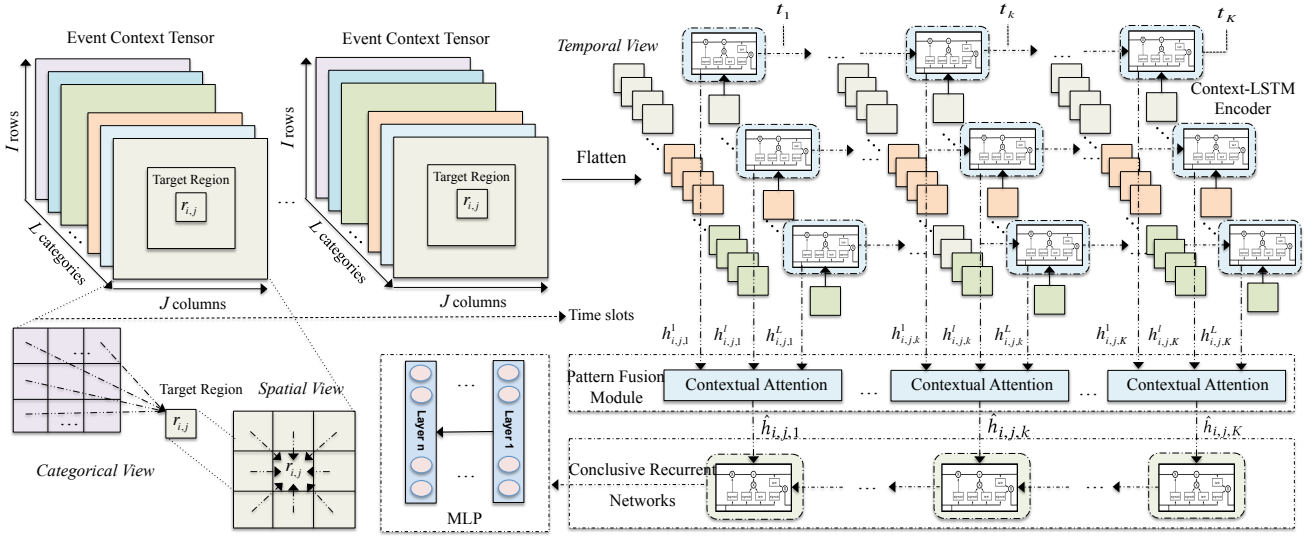


Figure 1: The Multiview and Multimodal Spatial-Temporal Learning Framework.

Multi-Modal Pattern Fusion Module. To jointly capture the inter-region and cross-category correlations in terms of abnormal event distributions, we propose a deep fusion module to model inherent occurrence patterns of abnormal events across surrounding geographical regions and different categories simultaneously. Given a specific time slot t_k, we supply the learned hidden representation vectors of each time-ordered sequence of tensor $\mathcal{A}^k_{i,j}$ to an attention mechanism in order to generate summarized embedding vectors from spatial-categorical views.

Conclusive Recurrent Networks. Relying on the hidden representations generated from spatial-temporal-categorical views, we then develop a conclusive recurrent networks to effectively capture the sequential patterns of cross-modal correlations between location, time, and event category. The final multi-view sequence representations from spatial-temporal-categorical dimensions are stored in the last states of the conclusive recurrent network cells, which provides a guidance for the prediction during the occurrence probability decoding phase.

3 METHODOLOGY

In this section, we present the details of our MiST framework.

3.1 Context-aware Recurrent Framework

In the architecture of MiST, we first employ Long Short-Term Memory (LSTM) networks to encode the complex intra-region correlations with respect to the distributions of abnormal events across time slots (from t_1 to t_K). In particular, LSTM consists of one cell state, and three controlled gates to update the cell state by performing write, read and reset operations, respectively. Formally, the updating functions for hidden states $\mathbf{h}^t_{i,j,l}$ and $\mathbf{c}^t_{i,j,l}$ corresponding to region $r_{i,j}$ and category c_l in t-th time step of input encoded sequence with length T in our LSTM encoder are presented as:

$$
\begin{aligned}
\mathbf{i}^t_{i,j,l} &= \sigma(\mathbf{W}_i \cdot \mathbf{h}^{t-1}_{i,j,l} + \mathbf{V}_i \cdot x^t_{i,j,l} + b_i) \\
o^t_{i,j,l} &= \sigma(\mathbf{W}_o \cdot \mathbf{h}^{t-1}_{i,j,l} + \mathbf{V}_o \cdot x^t_{i,j,l} + \mathbf{b}_o) \\
\mathbf{f}^t_{i,j,l} &= \sigma(\mathbf{W}_f \cdot h^{t-1}_{i,j,l} + V_f \cdot x^t_{i,j,l} + b_f) \\
\widetilde{c^t_{i,j,l}} &= \phi(\mathbf{W}_c \cdot \mathbf{h}^{t-1}_{i,j,l} + \mathbf{V}_c \cdot x^t_{i,j,l} + \mathbf{b}_c) \\
\mathbf{c}^t_{i,j,l} &= \mathbf{f}^t_{i,j,l} \odot \mathbf{c}^{t-1}_{i,j,l} + \mathbf{i}^t_{i,j,l} \odot \widetilde{\mathbf{c}^t_{i,j,l}} \\
\mathbf{h}^k_{i,j,l} &= \mathbf{o}^t_{i,j,l} \odot \phi(\mathbf{c}^t_{i,j,l})
\end{aligned} \tag{1}
$$

where $\mathbf{W}_* \in \mathbb{R}^{d_s \times d_s}$ represents the transformation matrix from the previous states (*i.e.*, $\mathbf{c}^{t-1}_{i,j,l}$ and $\mathbf{h}^{t-1}_{i,j,l}$) to LSTM cell and $\mathbf{V}_* \in \mathbb{R}^{d_x \times d_s}$ are the transformation matrices from input to LSTM cell. Here, d_x and d_s denotes the dimension of input vectors and hidden states, respectively. Furthermore, $\mathbf{b}_* \in \mathbb{R}^{d_s}$ is defined as a vector of bias term. $\sigma(\cdot)$ and $\phi(\cdot)$ represents the sigmoid and tanh function, respectively. The $\odot$ operator denotes the element-wise product. We denote the input gate, output gate and forget gate as $\mathbf{i}^t_{i,j,l}$, $\mathbf{o}^t_{i,j,l}$, and $\mathbf{f}^t_{i,j,l}$, respectively. For simplicity, we denote Eq. 1 as $\mathbf{h}^t_{i,j,l} = \text{LSTM}(*, \mathbf{c}^{t-1}_{i,j,l}, \mathbf{h}^{t-1}_{i,j,l})$ in the following subsections. While there exist other variations of recurrent neural networks, *e.g.*, gated recurrent unit (GRU), we choose LSTM for its general expressions.

3.2 Multi-Modal Pattern Fusion Module

While directly applying the recurrent neural network to solve the abnormal event forecasting problem is intuitive, the general recurrent neural network can do little to handle the complicated influencing factors from other geographical regions and event categories. Hence, we further employ an attention mechanism to adaptively capture the dynamic correlations from spatial and categorical view. Attention mechanisms are proposed to infer the importance of different parts of the training data, and let the learning algorithms focus on the most informative parts. Attention mechanism aims

to free the encoder-decoder architecture from the fixed-length internal representation by introducing a context vector to model the relevance. In addition, in order to differentiate regions and categories in the fusion process, we incorporate region embedding $\mathbf{e}_{r_{i,j}} \in \mathbb{R}^{d_e}$ and category embedding $\mathbf{e}_{c_j} \in \mathbb{R}^{d_e}$ as contextual information into the attention mechanism. Formally, the attention computation formulations in MiST are given:

$$\boldsymbol{\eta}^k_{i,j,l} = \tanh(\mathbf{W}^k(\mathbf{h}^k_{i,j,l}, \mathbf{e}_{r_{i,j}}, \mathbf{e}_{c_j}) + \mathbf{b}^k)$$
$$\alpha^k_{i,j,l} = \frac{\exp(\boldsymbol{\eta}^k_{i,j,l})}{\sum_{i,j\in G}\sum_{l=1}^{L}\exp(\boldsymbol{\eta}^k_{i,j,l})} \tag{2}$$

where we term the size of hidden representation vector in attention networks as *attention dimensionality* which is denoted as S. $\mathbf{W}^k \in \mathbb{R}^{d_s \times S}$ and $\mathbf{b}^k) \in \mathbb{R}^{d_s}$ are respectively the weight matrix and bias vector that project the input into hidden layers to get $\eta^k_{i,j,l}$ as hidden representation of $\mathbf{h}^k_{i,j,l}$. Then, we measure the importance of hidden representations of each region $r_{i,j}$ in the grid map $G = I \times J$ ($(i,j) \in G$) and category c_l ($c_l \in C$), and get a normalized importance weight $\alpha^k_{i,j,l}$ through a softmax function. The attention weights of our fusion module are jointly determined by the input spatial-categorical features and the encoded historical hidden states in the context-LSTM encoder. After obtaining the attention weights, the output hidden representation vector at time slot t_k is derived as:

$$\mathbf{q}^k = \sum_{i,j\in G}\sum_{l=1}^{L}\alpha^k_{i,j,l}\mathbf{h}^k_{i,j,l} \tag{3}$$

where $\mathbf{q}^k$ is the summarized concatenation of view-specific representations $\mathbf{h}^k_{i,j,l}$, describing what kind of influential factors will be considered as more important for abnormal event occurrences at target region $r_{i,j}$. In the training process of MiST, the developed the deep fusion module with attention mechanism is parametrized as a feed-forward neural network which can be trained with the whole neural network. Our proposed approach is very general, which can automatically learn the correlation weights of different views.

3.3 Conclusive Recurrent Network

So far we have developed two components of MiST to (i) model dynamic intra-region correlations from temporal views with the context-LSTM encoder; (ii) capture complex inter-region and cross-categorical correlations from spatial-categorical views with deep fusion module. With the above learned weights as coefficients, the summarized representation $\mathbf{q}^k$ are calculated as the weighted combinations of the view-specific representations with the voting weights (*i.e.*, $\alpha^k_{i,j,l}$) of different views.

To integrate the complex patterns encoded from spatial-categorical factors with the temporal patterns under the MiST framework, we propose to encode the multi-dimensional patterns with recurrent neural networks and model the complex location-time-category interactions with latent-space representations. In this paper, we apply LSTM as the recurrent unit and formulate it as follows:

$$\boldsymbol{\xi}_k = \text{LSTM}(\mathbf{q}_{k-1}, \boldsymbol{\xi}_{k-1}) \tag{4}$$

The joint embedding $\boldsymbol{\xi}$ maps all spatial, temporal, and categorical units into a common latent space. The developed conclusive recurrent networks provide a flexible way to let different views to collaborate with each other. By combining this spatial and categorical contextual signals with the current recurrent temporal status, our MiST framework could predict future abnormal events based on not only the sequential relation but also the spatial correlations across regions and co-occurrence relationships across categories.

3.4 Forecasting and Model Inference

Finally, we utilize the Multilayer Perceptron (MLP) component to decode the occurrence probability by capturing the non-linear dependencies between elements in hidden vectors. Formally, we represent MLP as follows:

$$\psi_1 = \phi(\mathbf{W}_1 \cdot \lambda_1 + \mathbf{b}_1)$$
$$\cdots\cdots$$
$$\psi_N = \phi(\mathbf{W}_n \cdot \lambda_n + \mathbf{b}_n)$$
$$y^{l,k}_{i,j} = \sigma(\mathbf{W}' \cdot L_n + \mathbf{b}') \tag{5}$$

where N represents the number of hidden layers in MLP (indexed by n). For the ψ_n layer, $\mathbf{W}_n$ and $\mathbf{b}_n$ represents the weight matrix and bias vector, respectively. We use *ReLU* (denoted as $\phi(\cdot)$) as the activation function of the fully connected layers and further specify the activation function as *sigmod* (denoted as $\sigma(\cdot)$) to output the abnormal event occurrence probability of category c_l at region $r_{i,j}$ in time slot t_k, *i.e.*, $y^{l,k}_{i,j}$.

In general, our abnormal event occurrence forecasting can be regarded as a classification problem. We utilize cross entropy as the metric in our loss function which is defined as follows:

$$\mathscr{L} = -\sum_{(i,j,l,k)\in S} y^{l,k}_{i,j}\log\hat{y}^{l,k}_{i,j} + (1 - y^{l,k}_{i,j})\log(1 - \hat{y}^{l,k}_{i,j}) \tag{6}$$

where $\hat{y}^{l,k}_{i,j}$ denotes the estimated probability of the l-th category of abnormal events occur at region $r_{i,j}$ in k-th time slot. Here, S is the set of abnormal events in the training process. The model parameters can be derived by minimizing the loss function. In this work, we use Adam optimizer [16] to learn the parameters of MiST.

4 EVALUATION

In this section, we conduct extensive experiments on three real-world abnormal event datasets (*i.e.*, crime and urban anomaly data) collected from New York City (NYC) and Chicago to verify the effectiveness of our developed *MiST* framework. We then analyze the experimental results and demonstrate the accuracy promotion by comparing it with various baselines. We perform experiments to answer the following questions:

- **Q1**: Compared with the state-of-the-art forecasting techniques, can our *MiST* framework achieve comparable accuracy in predicting both citywide crimes and anomalies across different cities?
- **Q2**: Does *MiST* consistently outperform other algorithms with respect to different time periods?

Algorithm 1: The Learning Process of MiST Model.

Input: Historical Abnormal Event Sequences Y, Neighbor Region Set G, Category Set L, Sequence Length T, and Batch Size b_{size}.

Paras: Embedding Matrices $e_r \in \mathbb{R}^{(I*J)\times d_e}$, $e_c \in \mathbb{R}^{C\times d_e}$, and Other Hidden Parameters θ.

Initialize all parameters;

// Sample a minibatch of size b_{size}.

foreach $\mathsf{T}_{\text{batch}} = \text{sample}(X, b_{\text{size}})$ **do**

  foreach $\langle i, j, l, t\rangle \in \mathsf{T}_{\text{batch}}$ **do**

    foreach $\langle i', j'\rangle \in G[(i, j)]$ **do**

      $R_{i,j} = e_r[(i, j) :]$;

      // Lookup region embeddings

      foreach $l' \in L$ **do**

        $C_l = e_c[l, :]$;

        // Lookup category embeddings

        $c^0_{i',j',l'} = c^0, h^0_{i',j',l'} = h^0$// init. hidden states

        for $t' \leftarrow (k - s)$ **to** $(k - 1)$ **do**

          $c^t_{i',j',l'}, h^t_{i',j',l'} = \text{LSTM}(X[i', j', l', t'], c^{t-1}_{i',j',l'}, h^{t-1}_{i',j',l'})$;

          // Eq.1

        end

      end

    end

    $\hat{T}_k = \text{Attention}(h^k_{:,:,:}, [h^k_{:,:,:}; R_{:,:}; C_{:}])$ $\hat{y}^t_{i,j,l} = \text{MLP}(\hat{T}_k)$

    $y^t_{i,j,l} = X[i, j, l, t]$ Update loss $\mathcal{L}$

  end

  Update all parameters w.r.t $\mathcal{L}$;

end

- **Q3**: How does our *MiST* model work for forecasting abnormal events with different categories compared with state-of-the-art techniques?
- **Q4**: How is the performance of *MiST* variants with different combinations of key components in the joint framework?
- **Q5**: How is the performance of *MiST* with different spatial and temporal scales?
- **Q6**: How do different hyperparameter settings affect the forecasting performance of *MiST*?
- **Q7**: How is the interpretation of our *MiST* framework in capturing dynamic importance weights across spatial and category dimensions when predicting citywide abnormal events?

4.1 Data Description

4.1.1 **Data Statistics.** We collect three abnormal event datasets with two different types from NYC and Chicago, *i.e.*, two crime data and one urban anomaly data, and separately perform experiments of predicting the occurrences of each categorical urban crimes and anomalies at each geographical region in a city. The basic statistics of the three datasets are shown in Table 1. In our experiments, we focus on several key categories and consider other categories as external type. We also show the geographical distributions of abnormal events with respect to different categories and time periods in Figure 2. Their details are introduced as follows:

Table 1: Dataset Statistics.

Data Source	**Crime Reports from NYC**	
Time Span	From Jan, 2015 to Dec, 2015	
Category	Burglary	Robbery
Number of Instances	14,967	16,886
Category	Felony Assault	Grand Larceny
Number of Instances	20,189	41,873
Data Source	**Anomaly Reports from NYC**	
Time Span	From Jan, 2014 to Dec, 2014	
Category	Noise	Illegal Parking
Number of Instances	134,690	57,374
Category	Blocked Driveway	Building/Use
Number of Instances	74,698	24,319
Data Source	**Crime Reports from Chicago**	
Time Span	From Jan, 2015 to Dec, 2015	
Category	Burglary	Robbery
Number of Instances	13,103	9,633
Category	Narcotics	Theft
Number of Instances	21,607	56,695
Category	Assault	Battery
Number of Instances	16,992	48,824
Category	Criminal Damage	Deceptive Practice
Number of Instances	28,590	14,084

- **NYC Crime Data (NYC-C)**. This crime dataset records multiple categories of crime reports (*e.g.*, burglary, robbery) in NYC and each crime record is in the format of (crime category, latitude, longitude, timestamp). This data was collected from Jan 2015 to Dec 2015.
- **NYC Urban Anomaly Data (NYC-A)**. This dataset spanning from Jan 2014 to Dec, 2014 is collected from 311 non-emergency services in New York City which records urban anomalies of different categories (*i.e.*, noise, illegal parking). Each reported anomaly record is formatted as (anomaly category, latitude, longitude, timestamp).
- **Chicago Crime Data (CHI-C)**. This crime data was collected from Chicago during 2015 to Dec 2015, which documents different categorical crime reports (*e.g.*, theft, narcotics). The data format is similar as that of NYC crime data.

4.2 Experimental Settings

4.2.1 **Parameter Settings.** In our experiments, we leverage Adam as the optimizer in the learning process of *MiST* and implemented *MiST* with TensorFlow architecture. We set the hidden state dimensionality d_s as 32 and embedding size d_e as 32. The attention dimension and the number of MLP layers is set to 32 and 3, respectively. Additionally, the batch size and learning rate is set to 64 and 0.001, respectively.

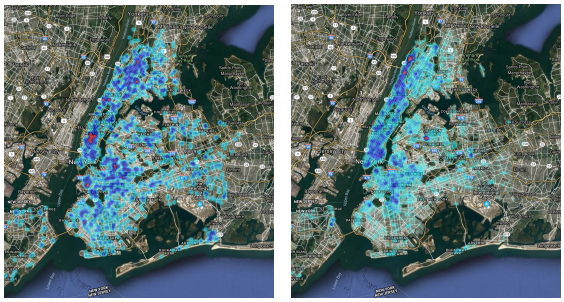
Table 2: Crime forecasting results across different categories in terms of *Macro-F1* and *Micro-F1*.

City	Month	Metrics	SVR	ARIMA	LR	ST-RNN	GRU	RDN	HRN	ARM	*MiST*
NYC-C	Aug	Macro-F1	0.3960	0.4090	0.4325	0.5070	0.5099	0.5102	0.4893	0.5031	**0.5384**
		Micro-F1	0.3010	0.3210	0.4108	0.4530	0.4605	0.4620	0.4650	0.3298	**0.5050**
	Sep	Macro-F1	0.4160	0.4190	0.4526	0.4891	0.4910	0.5015	0.4856	0.4868	**0.5365**
		Micro-F1	0.3210	0.3370	0.3881	0.4415	0.4439	0.4582	0.4627	0.3099	**0.5054**
	Oct	Macro-F1	0.4180	0.4190	0.4345	0.4903	0.4906	0.4907	0.4748	0.4907	**0.5110**
		Micro-F1	0.3210	0.3300	0.3727	0.4340	0.4330	0.4444	0.4451	0.3046	**0.4739**
	Nov	Macro-F1	0.4280	0.4290	0.4246	0.4705	0.4719	0.4896	0.4945	0.4908	**0.5066**
		Micro-F1	0.3280	0.3430	0.3768	0.4045	0.4370	0.4389	0.4478	0.3017	**0.4772**
	Dec	Macro-F1	0.4240	0.4150	0.3875	0.4683	0.4805	0.4806	0.4639	0.4815	**0.4920**
		Micro-F1	0.3320	0.3320	0.3450	0.3977	0.4147	0.4303	0.4117	0.3001	**0.4563**
CHI-C	Aug	Macro-F1	0.5004	0.5121	0.4838	0.5349	0.5189	0.5278	0.5391	0.5382	**0.5896**
		Micro-F1	0.3108	0.3302	0.3812	0.3949	0.4203	0.3728	0.3886	0.3871	**0.4663**
	Sep	Macro-F1	0.5159	0.5164	0.4733	0.4898	0.4796	0.5145	0.5095	0.5037	**0.5637**
		Micro-F1	0.3293	0.3412	0.3913	0.4292	0.4227	0.3551	0.3487	0.3408	**0.4649**
	Oct	Macro-F1	0.5141	0.5201	0.4671	0.4622	0.5341	0.5292	0.5354	0.5290	**0.5793**
		Micro-F1	0.3275	0.3432	0.2948	0.3665	0.4007	0.3781	0.3882	0.3817	**0.4695**
	Nov	Macro-F1	0.5111	0.5113	0.4941	0.5383	0.5181	0.5270	0.5330	0.5283	**0.5634**
		Micro-F1	0.3314	0.3417	0.3819	0.4168	0.3795	0.3877	0.3971	0.3764	**0.4564**
	Dec	Macro-F1	0.5034	0.5048	0.4802	0.5337	0.5221	0.5338	0.5103	0.5285	**0.5805**
		Micro-F1	0.3271	0.3343	0.3528	0.4053	0.3993	0.4067	0.3682	0.3987	**0.4783**
NYC-A	Aug	Macro-F1	0.6181	0.6072	0.5788	0.6455	0.6008	0.6496	0.6561	0.6511	**0.7119**
		Micro-F1	0.5405	0.5192	0.5614	0.5925	0.5441	0.5937	0.5942	0.5752	**0.6802**
	Sep	Macro-F1	0.6177	0.6204	0.5681	0.6797	0.6484	0.6733	0.6820	0.6670	**0.7244**
		Micro-F1	0.5353	0.5302	0.5268	0.6154	0.6115	0.6051	0.6032	0.5912	**0.6773**
	Oct	Macro-F1	0.6294	0.6303	0.5901	0.6625	0.6424	0.6711	0.6812	0.6672	**0.7278**
		Micro-F1	0.5436	0.5412	0.5385	0.5957	0.6072	0.6013	0.6173	0.5900	**0.6979**
	Nov	Macro-F1	0.6136	0.6214	0.5915	0.6559	0.6423	0.6541	0.6690	0.6463	**0.7034**
		Micro-F1	0.5349	0.5452	0.5670	0.5851	0.5891	0.5813	0.6007	0.5654	**0.6663**
	Dec	Macro-F1	0.5947	0.5849	0.5556	0.6135	0.6231	0.6178	0.6370	0.6247	**0.6895**
		Micro-F1	0.5126	0.5234	0.5212	0.5265	0.5495	0.5323	0.5183	0.5218	**0.6309**

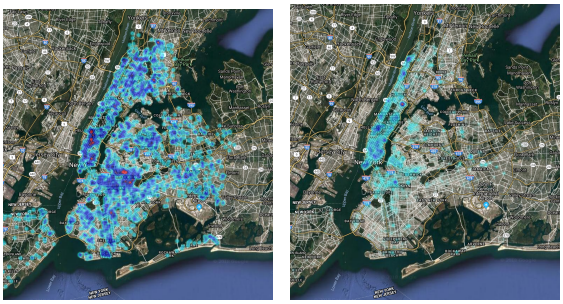
4.2.2 **Baseline Methods.** *MiST* is measured against the following baseline methods from five categories: (i) the conventional time series forecasting methods (*i.e.*, SVR and ARIMA); (ii) conventional supervised learning algorithm (*i.e.*, LR); (iii) recurrent neural network and its variants for spatial-temporal data forecasting (*i.e.*, ST-RNN and GRU); (iv) advanced neural network models for time series and sequence modeling (*i.e.*, RDN, HRN and ARM). The details are presented as follows:

- **Support Vector Regression (SVR)** [3]: it is a supervised learning model for time series regression analysis based on kernel functions which are characterized by the margin and the number of support vectors.
- **Auto-Regressive Integrated Moving Average (ARIMA)** [15]: it is a conventional time series prediction model for understanding and predicting future values in a time series which is in conjunction with stationary and linear transformations.
- **Logistic Regression (LR)** [11]: it is a statistical model which forecasts each region's abnormal event occurrences based on temporal features (e.g., the day of a week and the month of a year) extracted from historical crime logs.
- **Spatial-Temporal Recurrent Neural Network (ST-RNN)**[25]: it utilizes the recurrent neural networks to capture the spatial and periodical temporal contexts in modeling time-ordered data from each time intervals.
- **Gated Recurrent Framework (GRU)** [6]: it is a time-ordered sequence modeling approach based on gated recurrent unit and encodes the input time series into latent representations.
- **Recurrent Deep Networks (RDN)** [12]: this method is a recurrent deep network framework to consider long-term dependencies in recurrent neural networks, and aims to predict spatial events via jointly modeling normal dynamic temporal patterns.
- **Attentive Hierarchical Recurrent Networks (HRN)** [14]: it is a attention-based hierarchical recurrent framework that is capable of capturing the dynamic sequential patterns and their inherent interrelationships with other ubiquitous data.
- **Attention-based Recurrent Model (ARM)** [9]: this method leverages recurrent framework and attention mechanisms to interpret the relations between the current and past values in predicting spatial-temporal mobility data.

4.2.3 **Evaluation Protocols.** In our experiments, we chronologically split the dataset into training (6.5 months), validation (0.5 month), and test (1 month) sets. The validation set was used for tuning hyper-parameters and the final performance comparison was conducted on the test set. We partitioned New York City and



(a) Crime-Burglary-Aug (b) Anomaly-Noise-Aug



(c) Crime-Burglary-Nov (d) Anomaly-Noise-Nov

Figure 2: Distributions of abnormal events in NYC.

Chicago into 248 and 189 disjointed geographical regions, respectively, where each region is a $2km \times 2km$ grid, following the similar settings in [32, 43]. Based on the region partition results, we can map each abnormal event (*i.e.*, report of crime or urban anomaly) into an individual geographical region and generate the inputs for *MiST* framework. In this work, We adopted two types of evaluation metrics to fully evaluate all methods:

- (i) We use *Marco-F1* and *Micro-F1* [10, 33] to evaluate the prediction accuracy across different crime categories. These metrics indicate the overall performance across different classes. The mathematical definitions of Marco-F1 and Micro-F1 are given: Micro-F1 $= \frac{1}{J} \cdot \sum_{j=1}^{J} \frac{2 \cdot TP_j}{2 \cdot TP_j + FN_j + FP_j}$ and Macro-F1 $= \frac{2 \cdot \sum_{j=1}^{J} TP_j}{2 \cdot \sum_{j=1}^{J} TP_j + \sum_{j=1}^{J} FN_j + \sum_{j=1}^{J} FP_j}$. where J is the total number of event categories. A higher Micro-F1 and Macro-F1 score indicates better performance.
- (ii) we use *F1-score* (harmonic mean to balance precision and recall) and *AUC* [8] to evaluate the accuracy in predicting an individual category of abnormal event occurrence. A higher F1-score and AUC value reflects a better forecasting accuracy.

To ensure the fairness of performance comparison for all compared methods, our experiments are conducted by forecasting occurrences of abnormal events in consecutive days in the time period of test set. In the evaluation results, the average performance is reported across all days in the test time period.

4.3 Performance Comparison (Q1, Q2 and Q3)

4.3.1 **Overall Comparison (Q1).** Table 2 shows both the crime and urban anomaly forecasting accuracy across categories on different cities in terms of Macro-F1 and Micro-F1. From the evaluation results, we summarize three key observations as follows:

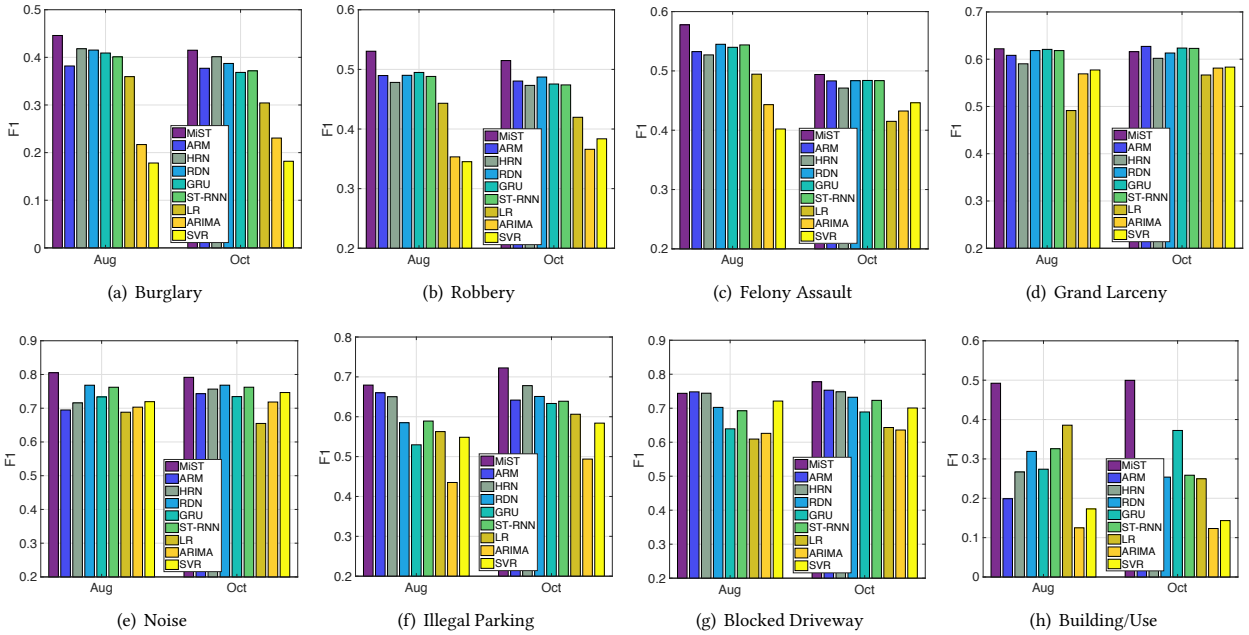
First and foremost, *MiST* is significantly better than other different types of neural network-based methods. For example, for predicting crimes in Chicago, *MiST* achieves relatively 9.6% and 30.9% improvements over the best performed baseline (*i.e.*, RDN) in terms of Macro-F1 and Micro-F1 on September. This sheds lights on the benefit of our multi-view model which jointly considers the spatial, temporal, and semantic relationships.

Second, the performance of neural network-based methods are superior to that of conventional time series forecasting techniques and supervised learning approach. This is due to the factors that: (1) conventional time series forecasting methods only emphasize one fixed temporal pattern, rather than the time-evolving temporal dependencies; (2) the neural network-based methods could capture complex inherent structures of multi-dimensional spatial-temporal data in a non-linear manner, which is more advantageous.

Lastly, there is no obvious winner among the recurrent network network variants (*i.e.*, ST-RNN and GRU) and deep sequential data modeling solutions (*i.e.*, RDN, HRN and ARM). This again confirms that only considering the data dependencies from temporal dimension is insufficient to predict both crime and urban anomaly occurrence well. In contrast, *MiST* dynamically associates weights for latent spatial, temporal and categorical correlations, which shows remarkable flexibility and superiority.

4.3.2 **Forecasting Accuracy v.s. Time Periods (Q2).** To make thorough evaluation, we conduct comparison experiment of *MiST* and all baselines across different training and test time periods. We can note that the best performance is consistently achieved by *MiST* with different forecasting time periods. In addition, another observation is that the performance gain between *MiST* and other baselines is relatively stable when sliding the training and testing time windows, which reflects the robustness of our model in learning the dynamic abnormal event distributions over time.

4.3.3 **Forecasting Accuracy v.s. Categories (Q3).** Furthermore, we investigate the effectiveness of *MiST* in predicting each individual category of abnormal events and report the evaluation results in Figure 4 and Figure 3 for crimes in Chicago and crimes as well as urban anomalies in NYC. We can observe that our *MiST* achieves the best performance in all cases. For example, *MiST* outperforms the best performed baseline (*i.e.*, RDN) by 13.2% on average in forecasting burglaries and 12.4% on average in forecasting illegal parking in NYC. Additionally, another interesting observation is that obvious improvements can also be obtained by *MiST* in predicting abnormal events of sparser categories, *e.g.*, 84.1% average relatively improvement over the state-of-the-art approach ST-RNN for building/use prediction. This observation indicates that *MiST* is capable of effectively modeling latent correlations from different contextual modalities (region, time, category), to alleviate the data scarcity problem in forecasting sparse anomalies.



(a) Burglary (b) Robbery (c) Felony Assault (d) Grand Larceny

(e) Noise (f) Illegal Parking (g) Blocked Driveway (h) Building/Use

Figure 3: Forecasting results for individual category in terms of *F1-score* on NYC crime and anomaly data.

4.4 Component-Wise Evaluation of *MiST* (Q4)

In addition to comparing *MiST* with state-of-the-art techniques, we also aim to get a better understanding of key components of *MiST* by studying the effect of different view components. We denote the full version of our method as as *MiST*.

- **Spatial-View + Temporal View** *MiST*-st: This variant captures both spatial and temporal dependencies in predicting abnormal events, *i.e.*, does not consider cross-categorical influences.
- **Category-View + Temporal View** *MiST*-ct: This variant considers both categorical and temporal views, *i.e.*, without considering inter-region spatial correlations.
- **Temporal View** *MiST*-t: For this variant, we only use LSTM with temporal attention mechanism to consider the temporal dimension of muli-view data, *i.e.*, without considering contextual signals from both spatial and categorical views.

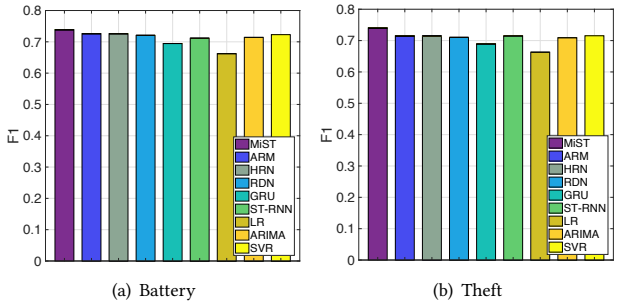
We report the evaluation results on crime dataset in Figure 5. We can notice that the full version of our developed framework *MiST* achieves the best performance in all cases. As such, it is necessary to build a joint framework to capture spatial view (inter-region spatial correlations), temporal view (intra-region temporal correlations), as well as categorical view (cross-categorical dependencies) simultaneously for forecasting citywide abnormal events.

4.5 Effect of Spatial and Temporal Scale (Q5)

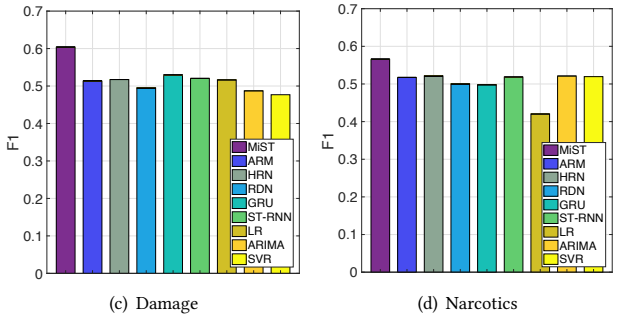
In our evaluations, we further investigated the effect of spatial and temporal scale in *MiST*, *i.e.*, the geographical scale of grid map $G = I \times J$ in the event context tensor $\mathcal{A}$ ($I = J$ in our experiments) and the length of encoded sequence T in our developed recurrent framework. The evaluation results are presented in Figure 6. There are two interesting insights from the result: (i) Figure 6(a) reveals that increasing I and J improves performance, which is not surprising since we consider a larger geographical space to learn jointly representations from mutli-views. Furthermore, as the spatial scale of the event context tensor increases, the performance slightly increases but mainly remains stable when $I = J = 11$. One potential reason is that when considering larger geographical coverage, more parameters need to be learned. As a result, the training of *MiST* becomes harder. (ii) As depicted in Figure 6(b), the performance of *MiST* becomes better as the encoded sequence length T increases and tends to saturate once T reaches 10. Due to space limit, we only report the results of crime data on Oct. Similar results can be observed on urban anomaly dataset and other time periods.

4.6 Hyperparameters Studies (Q6)

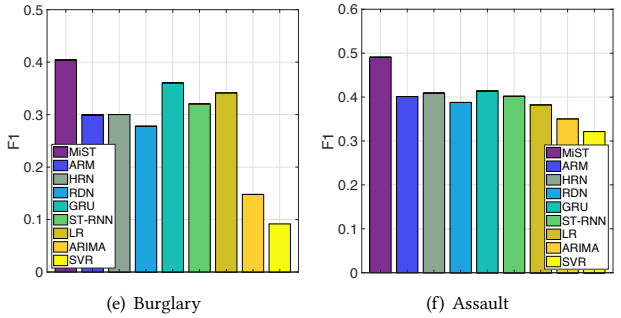
To investigate the robustness of *MiST* framework, we examine how the different choices of parameters affect the performance of *MiST* in the prediction performance. Except for the parameter being tested, we set other parameters at the default values. Figure 7 and Figure 8 shows the evaluation results as a function of one selected parameter when fixing others. Overall, we observe that *MiST* is not strictly sensitive to these parameters in two tasks and is able to reach high performance under a cost-effective parameter choice, which demonstrates the robustness of *MiST*. Furthermore, we can observe that the increase of prediction performance saturates as the representation dimensionality reaches around 32. This is because: at the beginning, a larger value of hidden state dimensionality d_s and attention dimension S brings a stronger representation power for the recent framework and attention network, but the further increase of dimension size of latent representations to encode abnormal event occurrence patterns from multi-views might lead to the overfitting issue. In our experiments, we set the dimension



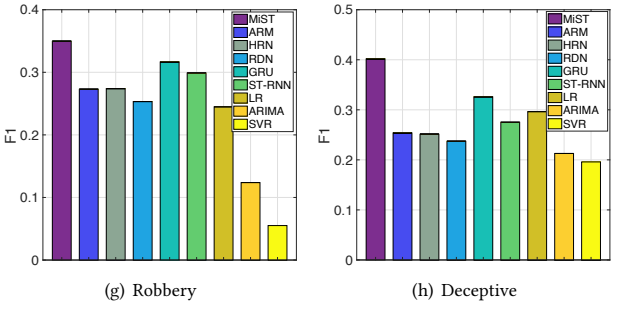
(a) Battery (b) Theft



(c) Damage (d) Narcotics

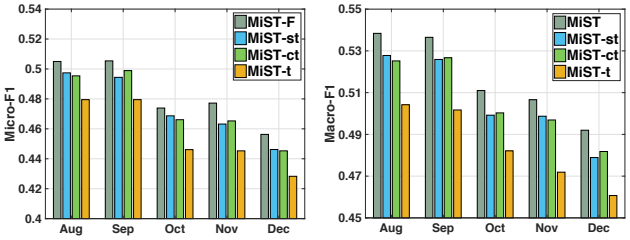


(e) Burglary (f) Assault



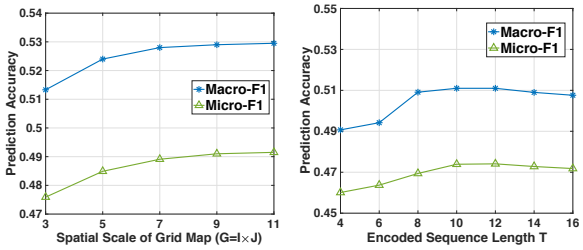
(g) Robbery (h) Deceptive

Figure 4: Forecasting results for individual category in terms of *F1-score* on Chicago crime data.



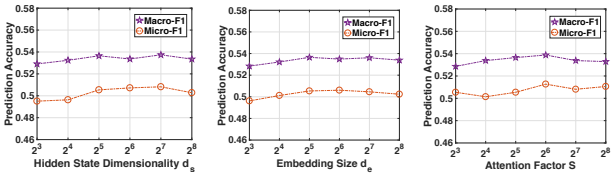
(a) Micro-F1 (b) Macro-F1

Figure 5: Evaluation on MiST variants.



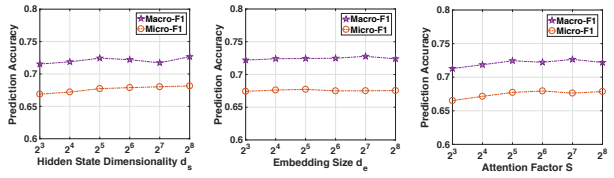
(a) Spatial Scale Effect (b) Temporal Scale Effect

Figure 6: Effect of spatial and temporal scale in *MiST* on crime data in terms of *Macro-F1* and *Micro-F1* on Oct.



(a) Hidden State Dimensionality d_s (b) Embedding Size d_e (c) Attention Dimension S

Figure 7: Parameter sensitivity of *MiST* in predicting crimes in NYC in terms of *Macro-F1* and *Micro-F1* on Sep.



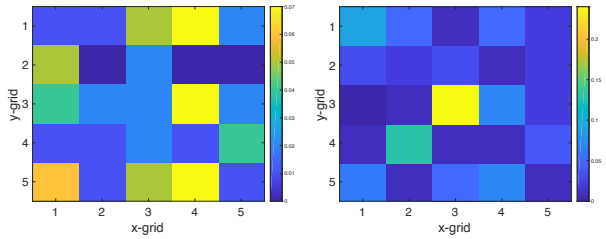
(a) Hidden State Dimensionality d_s (b) Embedding Size d_e (c) Attention Dimension S

Figure 8: Parameter sensitivity of *MiST* in predicting urban anomalies in NYC in terms of *Macro-F1* and *Macro-F1* on Sep.

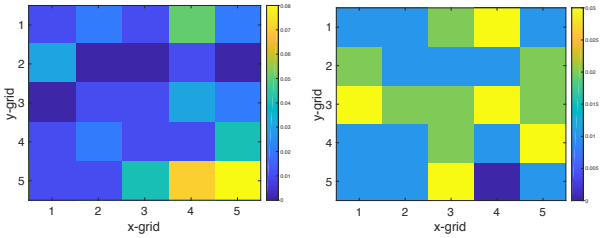
size as 32 due to the consideration of the trade-off between the effectiveness and computational cost.

4.7 Case Study (Q7)

Apart from the superior forecasting performance, another key advantage of *MiST* is its ability in interpreting the importance weights of spatial and categorical correlations in predicting a specific category of abnormal event at a region. To demonstrate this, we perform case studies to show the explainability of our framework by visualizing the attention weights obtained from *MiST* in predicting burglaries with a 5×5 grid map (the target region is in the middle of this map) in NYC, as shown in Figure 9. We could observe that



(a) Robbery (b) Burglary



(c) Felony Assault (d) Grand Larceny

Figure 9: The importance weights of across regions and crime categories learned from *MiST* in predicting burglaries with a 5×5 grid map in NYC.

MiST enables the dynamic modeling of correlations between the target region and other regions, as well as latent influences between the target event category (burglary) and other categories.

5 RELATED WORK

We introduce the research work which are related to our study.

Ubiquitous Sensing Applications. Ubiquitous sensing has emerged as a new sensing paradigm that empowers average people to contribute their observations and measurements about the physical world. Numerous novel applications have been developed to address various challenges in smart cities applications [4, 18, 21, 32, 36, 37, 41, 42], such as intelligent transportation [4], privacy-preserving [37], business assessment [21] and human activity modeling [41]. However, the abnormal event prediction problem in ubiquitous sensing scenario remain as a challenging problem to be solved. In this paper, we develop an end-to-end model to tackle this problem by addressing the unique challenges.

Spatial Anomaly Detection and Event Forecasting. There exists a good amount of work on the topics *geographical anomaly detection* [7, 17, 44, 46, 47]. For example, Doan *et al.* detected anomalies by modelling the behaviours of pedestrian flows across multiple locations [7]. Zheng *et al.* identified the anomalies from spatial-temporal data using multiple data sources [47]. However, all above solutions identified the geographical anomalies *after they happen* which might lead to the inefficiency to handle the anomalies beforehand. Additionally, while several recent studies aim to predict social events using social media data [27, 31, 45], these work only consider obvious social events which can be identified based on the external content (*e.g.*, keywords) from social media platforms (*e.g.*, Twitter). In contrast, this paper solves a new problem of different categories of abnormal event prediction at each region of a city with a multi-view and multi-modal deep neural network architecture.

Sequential Data Modeling. A relevant body of work in the machine learning and data mining communities have investigated modeling sequential data [9, 12, 13, 35, 38, 48]. For example, Autoregression-based models (*e.g.*, ARIMA) have been proposed to study the problem of time series prediction [19]. Compared to traditional time series data, temporally-ordered abnormal event sequence has its own characteristics, *e.g.*, complex spatial-temporal correlations and dynamic multi-dimensional interactions. Recently, Recurrent neural networks (RNNs) become an effective tool due to their success in sequence modeling [1]. For example, Hu *et al.* proposed to capture sequential structures of subevents with LSTM networks [12]. Wu *et al.* developed a RNN based model to predict future behavioral trajectories of users in online recommendation platforms [35]. However, these works can only capture temporal dependency in time series, which ignore the various characteristics of geolocation data from multi-views. To overcome this problem, our MiST framework aims to forecast the abnormal event data in urban space via modeling latent correlations from spatial-temporal-categorical views.

Attention Mechanism. The attention mechanism has been proposed to strengthen not only the ability of RNN in capturing the long-range dependencies but also the interpretability in image analysis [5] and nature language processing [22]. Later, a number of studies developed attention-based models to encode hidden states in different time series applications, such as patients' health information prediction [28] and online personalized recommendation [2]. Inspired by the above work, this work proposes a deep fusion module based on attention mechanism. Different from the above attention-based methods which only focus on sequential patterns, we propose a novel multi-view attention model which fuses view-specific representations on different dimensions.

6 CONCLUSION

This paper explored a new neural network architecture MiST to explicitly model the dynamic patterns of citywide abnormal events from spatial-temporal-categorical views. We integrated recurrent neural networks with a multi-modal pattern fusion module to model spatial-temporal correlations. We evaluated our new MiST framework on different types of real-world datasets and the results showed that our approach achieves better performance when competing with several baselines. There are several future directions for our work. *First*, we are interested in extending the current method to identify the causalities for the occurrences of different categorical abnormal events, which is useful for suggesting relevant public policy making. For future work, it is promising to explore the factors underlying abnormal event occurrences and how different categorical event propagate with respect to spatio-temporal dimensions. *Second*, in this paper, we evaluated our method on three real-world datasets, *i.e.*, crime and anomaly data, due to limited data availability. In fact, the framework of MiST is general, and it is feasible to apply our model on other multi-dimensional time-stamped sequence data.

ACKNOWLEDGMENTS

This work is supported in part by the Army Research Laboratory under Cooperative Agreement Number W911NF- 09-2-0053 and the National Science Foundation (NSF) grants CNS- 1629914 and IIS-1447795. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

REFERENCES

[1] Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. 2015. Scheduled sampling for sequence prediction with recurrent neural networks. In *NIPS*. 1171–1179.

[2] Da Cao, Xiangnan He, Lianhai Miao, Yahui An, Chao Yang, and Richang Hong. 2018. Attentive Group Recommendation. In *Special Interest Group on Information Retrieval (SIGIR)*.

[3] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: a library for support vector machines. *Transactions on Intelligent Systems and Technology (TIST)* 2, 3 (2011), 27.

[4] Longbiao Chen, Daqing Zhang, Leye Wang, Dingqi Yang, et al. 2016. Dynamic cluster-based over-demand prediction in bike sharing systems. In *International Joint Conference on Pervasive and Ubiquitous Computing (Ubicomp)*. ACM, 841–852.

[5] Jongwon Choi, Hyung Jin Chang, Jiyeoup Jeong, Yiannis Demiris, and Jin Young Choi. 2016. Visual tracking using attention-modulated disintegration and integration. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 4321–4330.

[6] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, et al. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014).

[7] Minh Tuan Doan, Sutharshan Rajasegarar, Mahsa Salehi, Masud Moshtaghi, and Christopher Leckie. 2015. Profiling pedestrian distribution and anomaly detection in a dynamic environment. In *International Conference on Information and Knowledge Management (CIKM)*. ACM, 1827–1830.

[8] Yuxiao Dong, Jing Zhang, Jie Tang, Nitesh V Chawla, and Bai Wang. 2015. Coupledlp: Link prediction in coupled networks. In *International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 199–208.

[9] Jie Feng, Yong Li, Chao Zhang, and etc. 2018. DeepMove: Predicting Human Mobility with Attentional Recurrent Networks. In *International World Wide Web Conference (WWW)*. International World Wide Web Conferences Steering Committee, 1459–1468.

[10] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 855–864.

[11] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. 2013. *Applied logistic regression*. Vol. 398. John Wiley & Sons.

[12] Linmei Hu, Juanzi Li, Liqiang Nie, Xiao-Li Li, and Chao Shao. 2017. What Happens Next? Future Subevent Prediction Using Contextual Hierarchical LSTM.. In *International Conference on Artificial Intelligence (AAAI)*. 3450–3456.

[13] Chao Huang, Xian Wu, and Dong Wang. 2016. Crowdsourcing-based urban anomaly prediction system for smart cities. In *International on Conference on Information and Knowledge Management (CIKM)*. ACM, 1969–1972.

[14] Chao Huang, Junbo Zhang, Yu Zheng, and Nitesh V Chawla. 2018. DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction. In *International Conference on Information and Knowledge Management (CIKM)*. ACM, 1423–1432.

[15] Mehdi Khashei, Mehdi Bijari, and Gholam Ali Raissi Ardali. 2009. Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs). *Neurocomputing* 72, 4-6 (2009), 956–967.

[16] Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[17] Viet Duc Le, Hans Scholten, and Paul Havinga. 2013. Flead: Online frequency likelihood estimation anomaly detection for mobile sensing. In *International Joint Conference on Pervasive and Ubiquitous Computing (Ubicomp)*. ACM, 1159–1166.

[18] Ruirui Li, Jyun-Yu Jiang, Chelsea J-T Ju, and Wei Wang. 2019. CORALS: Who Are My Potential New Customers? Tapping into the Wisdom of Customers' Decisions. In *International Conference on Web Search and Data Mining (WSDM)*. ACM, 69–77.

[19] Xiaolong Li, Gang Pan, Zhaohui Wu, and etc. 2012. Prediction of urban human mobility using large-scale taxi traces and its applications. *Frontiers of Computer Science* 6, 1 (2012), 111–121.

[20] Defu Lian, Kai Zheng, Yong Ge, Longbing Cao, Enhong Chen, and Xing Xie. 2018. GeoMF++: scalable location recommendation via joint geographical modeling and matrix factorization. *Transactions on Information Systems (TOIS)* 36, 3 (2018), 33.

[21] Jianxun Lian, Fuzheng Zhang, Xing Xie, and Guangzhong Sun. 2017. Restaurant survival analysis with heterogeneous information. In *International World Wide Web Conference (WWW)*. ACM, 993–1002.

[22] Yankai Lin, Shiqi Shen, Zhiyuan Liu, Huanbo Luan, and Maosong Sun. 2016. Neural relation extraction with selective attention over instances. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, Vol. 1. 2124–2133.

[23] Christoph Ling, Jerome Kunegis, Sergej Sizov, and Steffen Staab. 2014. Detecting no-gaussian geographical topics in tagged photo collections. In *International Conference on Web Search and Data Mining (WSDM)*. ACM, 603–612.

[24] Fu-Yanjie Yao Zijun Xiong Hui Liu, Bin. 2013. Learning gepgrapihcal preferences for point-of-interest recommendation. In *International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 1043–1051.

[25] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. Predicting the Next Location: A Recurrent Model with Spatial and Temporal Contexts.. In *International Conference on Artificial Intelligence (AAAI)*. 194–200.

[26] Yu Liu, Baojian Zhou, Feng Chen, and David Cheung. 2017. Graph topic scan statistic for spatial event detection. In *International Conference on Information and Knowledge Management (CIKM)*. ACM, 489–498.

[27] Zhiwei Liu, Yang Yang, Zi Huang, and etc. 2017. Event early embedding: predicting event volume dynamics at early stage. In *Special Interest Group on Information Retrieval (SIGIR)*. ACM, 997–1000.

[28] Fenglong Ma, Radha Chitta, Jing Zhou, Quanzeng You, Tong Sun, and Jing Gao. 2017. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In *International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 1903–1911.

[29] Bei Pan, Ugur Demiryurek, and Cyrus Shahabi. 2012. Utilizing real-world transportation data for accurate traffic prediction. In *ICDM*. IEEE, 595–604.

[30] Goce Ristanoski, Wei Liu, and James Bailey. 2013. Time seriess forecasting using distribution enhanced linear regression. In *PAKDD*. Springer, 484–495.

[31] Minglai Shao, Jianxin Li, Feng Chen, and etc. 2017. An efficient approach to event detection and forecasting in dynamic multivariate social media networks. In *WWW*. ACM, 1631–1639.

[32] Hongjian Wang, Daniel Kifer, Corina Graif, and Zhenhui Li. 2016. Crime rate inference with big data. In *International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 635–644.

[33] Jinpeng Wang, Gao Cong, Wayne Xin Zhao, and Xiaoming Li. 2015. Mining User Intents in Twitter: A Semi-Supervised Approach to Inferring Intent Categories for Tweets.. In *International Conference on Artificial Intelligence (AAAI)*. 318–324.

[34] Hua Wei, Guanjie Zheng, Huaxiu Yao, and Zhenhui Li. 2018. IntelliLight: a reinforcement learning approach for intelligent traffic light control. In *International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM.

[35] Chao-Yuan Wu, Amr Ahmed, Alex Beutel, Alexander J Smola, and How Jing. 2017. Recurrent recommender networks. In *International Conference on Web Search and Data Mining (WSDM)*. ACM, 495–503.

[36] Xian Wu, Yuxiao Dong, Chao Huang, Jian Xu, Dong Wang, and Nitesh V Chawla. 2017. Uapd: Predicting urban anomalies from spatial-temporal data. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD)*. Springer, 622–638.

[37] Dingqi Yang, Daqing Zhang, Bingqing Qu, and Philippe Cudré-Mauroux. 2016. PrivCheck: privacy-preserving check-in data publishing for personalized location based services. In *International Joint Conference on Pervasive and Ubiquitous Computing (Ubicomp)*. ACM, 545–556.

[38] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye, and Zhenhui Li. 2018. Deep multi-view spatial-temporal network for taxi demand prediction. In *International Conference on Artificial Intelligence (AAAI)*.

[39] Rose Yu, Yaguang Li, Ugur Demiryurek, Cyrus Shahabi, and Yan Liu. 2017. Deep learning: a generic approach for extreme condition traffic forecasting. In *SIAM International Conference on Data Mining (SDM)*. SIAM, 777–785.

[40] Chao Zhang, Liyuan Liu, Dongming Lei, and etc. 2017. TrioVecEvent: embedding-based online local event detection in geo-tagged tweet streams. In *International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 595–604.

[41] Chao Zhang, Keyang Zhang, Quan Yuan, and etc. 2017. Regions, periods, activities: uncovering urban dynamics via cross-modal representation learning. In *International World Wide Web Conference (WWW)*. ACM, 361–370.

[42] Fuzheng Zhang, Nicholas Jing Yuan, David Wilkie, Yu Zheng, and Xing Xie. 2015. Sensing the pulse of urban refueling behavior: A perspective from taxi mobility. *Transactions on Intelligent Systems and Technology (TIST)* 6, 3 (2015), 37.

[43] Junbo Zhang, Yu Zheng, and Dengkang Qi. 2017. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *International Conference on Artificial Intelligence (AAAI)*.

[44] Xuchao Zhang, Liang Zhao, Arnold P Boedihardjo, Chang-Tien Lu, and Naren Ramakrishnan. 2017. Spatiotemporal event forecasting from incomplete hyper-local price data. In *International Conference on Information and Knowledge Management*

(CIKM). ACM, 507–516.

[45] Liang Zhao, Junxiang Wang, and Xiaojie Guo. 2018. Distant-supervision of heterogeneous multitask learning for social event forecasting with multilingual indicators. International Conference on Artificial Intelligence (AAAI).

[46] Guanjie Zheng, Susan Brantely, Thomas Lauvaus, and Li Zhenhui. 2017. Contextual spatial outlier detection with metric learning. In *International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 2161–2170.

[47] Yu Zheng, Huichu Zhang, and Yong Yu. 2015. Detecting collective anomalies from multiple spatio-temporal datasets across different domains. In *International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*. ACM.

[48] Xian Zhou, Yanyan Shen, Yanmin Zhu, and Linpeng Huang. 2018. Predicting multi-step citywide passenger demands using attention-based neural networks. In *International Conference on Web Search and Data Mining (WSDM)*. ACM, 736–744.