CityGuard: Citywide Fire Risk Forecasting Using A Machine Learning Approach

QIANRU WANG*, Northwestern Polytechnical University, China
JUNBO ZHANG†, JD Intelligent Cities Business Unit, JD Digits, Beijing, China and JD Intelligent Cities Research, China
BIN GUO†, Northwestern Polytechnical University, China
ZEXIA HAO and YIFANG ZHOU, JD Intelligent Cities Business Unit, JD Digits, Beijing, China and JD Intelligent Cities Research, China
JUNKAI SUN, Xidian University, China
ZHIWEN YU, Northwestern Polytechnical University, China
YU ZHENG, JD Intelligent Cities Business Unit, JD Digits, Beijing, China and JD Intelligent Cities Research, China

Forecasting the fire risk is of great importance to fire prevention deployments in a city, which can reduce loss even deaths caused by fires. However, it is very challenging because fires are influenced by many complex factors, including spatial correlations, temporal dependencies, even the mixture of these two and external factors. Firstly, the fire risk of a region is influenced by temporal effect of internal factors (*e.g.*, the historical fire risk records) and temporal effect of external factors (*e.g.*, weather). Secondly, a region's fire risk is not only influenced by its inherent geospatial attributes (*e.g.*, POIs) but also dependent on other regions in spatial. To address these challenges, we propose a machine learning approach to forecast the fire risk, entitled NeuroFire. NeuroFire can represent internal and external temporal effect then combine the temporal representation and spatial dependencies by a spatial-temporal loss function. Experimental evaluations on real-world datasets show that our NeuroFire outperforms 9 baselines, demonstrating the performance of our approach by several visualizations. Moreover, we implement a citywide fire forecasting system named CityGuard to display the analysis and forecasting results, which can assist the fire rescue department in deploying fire prevention.

CCS Concepts: • **Information systems → Information systems applications**; **Spatial-temporal systems**.

Additional Key Words and Phrases: Fire risk, Neural network, Conditional random field, Spatial-temporal data, Urban computing

*The research was done when the first author was an intern in JD Intelligent Cities Research.
†Corresponding author

Authors' addresses: Qianru Wang, Northwestern Polytechnical University, China; Junbo Zhang, msjunbozhang@outlook.com, JD Intelligent Cities Business Unit, JD Digits, Beijing, China and JD Intelligent Cities Research, China; Bin Guo, guob@nwpu.edu.cn, Northwestern Polytechnical University, China; Zexia Hao; Yifang Zhou, JD Intelligent Cities Business Unit, JD Digits, Beijing, China and JD Intelligent Cities Research, China; Junkai Sun, Xidian University, China; Zhiwen Yu, Northwestern Polytechnical University, China; Yu Zheng, JD Intelligent Cities Business Unit, JD Digits, Beijing, China and JD Intelligent Cities Research, China.

2474-9567/2019/12-ART156 $15.00
https://doi.org/10.1145/3369814

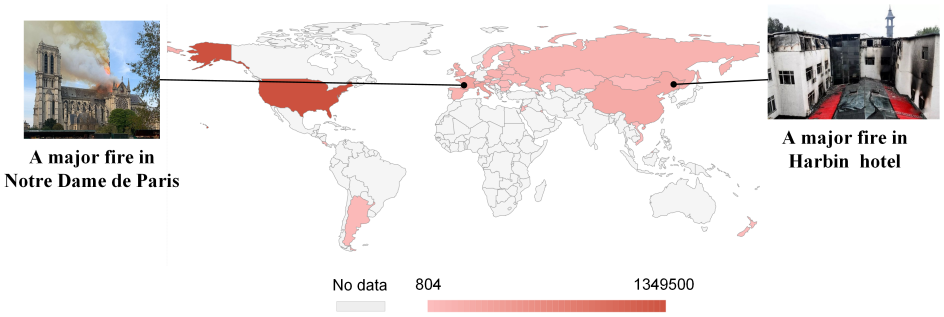


Fig. 1. Fire incidents distribution in the world. The picture is created with datamap (https://datamaps.co/).

ACM Reference Format:
Qianru Wang, Junbo Zhang, Bin Guo, Zexia Hao, Yifang Zhou, Junkai Sun, Zhiwen Yu, and Yu Zheng. 2019. CityGuard: Citywide Fire Risk Forecasting Using A Machine Learning Approach. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 4, Article 156 (December 2019), 21 pages. https://doi.org/10.1145/3369814

1 INTRODUCTION

Urban fires occur primarily in cities or towns, causing much financial loss, many injuries and even deaths. The International Association of Fire and Rescue Services (CTIF)[1], the largest firefighting organization in the world, reported that there were 3,115,061 fires across 34 countries in 2017 (as shown in Fig.1), which caused 47,948 injuries and 16,808 deaths. In particular, 53.2% of fires are structure and vehicle fires, which mainly happened in urban areas. For example, on April 15th 2019, a major fire engulfed the Notre-Dame de Paris[2] as shown in Fig.1. People from all over the world expressed sadness for the damage of treasures. As China urbanizes rapidly, fire and rescue departments face greater and greater challenges in reducing the risk that urban fires pose to public safety and residents' lives and property, all the more given their often limited resources. In 2018 alone, there were 237,000 fires, causing 1,407 civilian deaths, 798 civilian injuries and 3.675 billion RMB in property damage. For instance, on August 25th 2018, a fire broke out in a hotel in downtown Harbin[3] as shown in Fig.1, killing 20 people and injuring more than 20 people. Therefore, the threat that urban fires pose to public safety cannot be overlooked. To reduce the loss caused by fires, it is necessary to forecast the fire risk of each region in space, which we partition into grids. If fire and rescue departments can identify the regions with the highest fire risk, they can best deploy their limited resources to prevent urban fires.

As far as we know, there are few works about forecasting fire risk in urban areas. Existing work [22, 29] focused on predicting the fire risk without adding a temporal dimension to the risk prediction, which is less useful for fire and rescue departments. To meet this temporal requirement, we face some challenges because the fire risk of each grid (*e.g.*, dividing a city into 1km×1km) in the city is influenced by various factors:

(1) **Temporal effect.** The fire risk of a grid is influenced by *internal temporal effects* (*e.g.*, historical fire risk). the fire risk will be high if the historical fire risks are high. The fire risk is also correlated to the *external temporal effects* (*e.g.*, temperature, humidity, E-commerce orders). Some of external temporal effects have an immediate impact. For example, the fire risk will increase when the temperature is high or the humidity is low. And some of them have a delayed impact. For example, the fire risk will not increase immediately when

[1] www.ctif.org
[2] https://en.wikipedia.org/wiki/Notre-Dame_de_Paris_fire
[3] https://en.wikipedia.org/wiki/Harbin_hotel_fire

electronic orders increase, but will rise as more people uses these devices simultaneously. The analysis is detailed in Section 2.2. Therefore, we should integrate the internal temporal effects and external factors to find the relationship between temporal effects and fire risk.

(2) **Spatial factors.** Local spatial attributes and global spatial dependencies are both correlated to the fire risk of grids. *Local spatial attributes* (*e.g.*, POIs, human activities, population of area) infer the inherent attributes of the grid. For instance, fires are more likely to break out in a grid when the POI categories is diverse, which increase the crowd flow in the grid. Besides local spatial attributes, *global spatial dependencies* characterize the distribution of fire risk across the city. In the spatial dimension, some regions always have high fire risk, but the fire risks of some regions always stay low. The analysis is detailed in Section 2.2. To forecast the fire risk of each grid, we need to incorporate spatial attributes and dependencies into our model.

To address these challenges, we propose a machine learning approach named NeuroFire to forecast the fire risks of grids in two steps: temporal fire classification and spatial fire risk forecasting. In this paper, we define the fire risk as the probability of fires in a month. We leverage the GRU-CRF [15] (combining Gate Recurrent Unit [6] and Conditional Random Field [19]) model to represent the relationship of historical fire incidents and capture the immediate and delayed temporal impact of urban factors. We then combine the temporal representation and spatial dependencies of grids by designing a spatial-temporal loss function. The main contributions are as follows:

(1) We propose a citywide fire forecasting model named NeuroFire which utilizes GRU to learn temporal representations of urban data and integrate the temporal representations into fire risk sequences by CRF.
(2) We design a spatial-temporal loss function, which is used to model spatial dependencies between grids.
(3) We conduct experiments on real-world datasets collected from Zhengzhou, China. Evaluation results demonstrate that NeuroFire significantly outperforms 9 popular baselines. In addition, we develop a system named CityGuard to visualize the analysis and results.

2 DATASETS AND DATA ANALYSIS

2.1 Datasets

In this paper, we analyze and forecast the fire risk in Zhengzhou, China. We collected the following 4 real-world datasets from Jan. 2014 to Nov. 2018 as following:

- **Fire records:** We obtained fire records in Zhengzhou, Henan province from the Henan fire and rescue department. The dataset consists of location (latitude and longitude) and time of a fire. We divide Zhengzhou into grids. For example, there are 5472 grids when grids size is 1km×1km. Over 96% grids in city (5472 grids in total) are without fires in each month. For each grid, fires only occur in average of 2 months during 59 months.
- **POIs:** We obtained our POI dataset from OpenStreet Map[4], which contains the categories and locations of POIs.
- **Meteorology:** We crawled temperature and humidity from Weather China[5] for each day, and calculated average temperature and humidity for each month.
- **E-commerce orders:** E-commerce orders' information is indicative of fire risk. Firstly, electrical and mechanical failure is a main classification of fires [16]. Too many electronic products will result in overuse of electric power, which is one of major factors causing fires. The increasing number of electronic products will be found in the increasing E-commerce orders of electronic products. Secondly, orders' information is a useful feature to forecast fire risk. Existing work used parcels information as address-level features [29]. Our E-commerce dataset, which also has parcels information (orders information), contains the order

[4]https://www.openstreetmap.org/
[5]http://www.weather.com.cn/

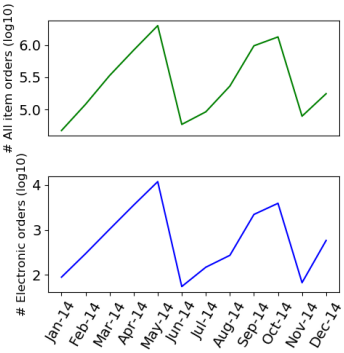


Fig. 2. Trend of all item orders and electronics orders

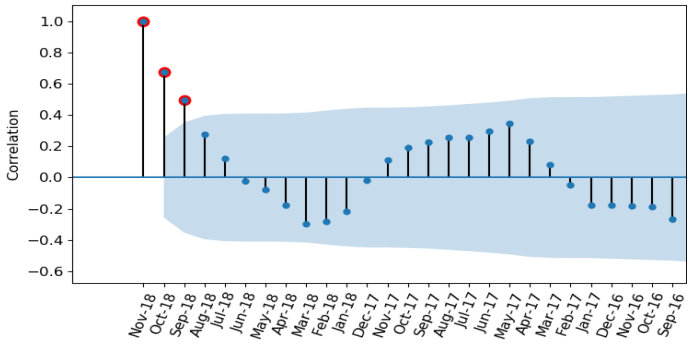


Fig. 3. Correlation between the current fire risk and historical fire risk

number of electronics (electronic orders) and the order number of all products in each grid for every month, which are correlated with fire incidents in temporal and spatial as following illustration in Section 2.2. Due to the lack of data on electronics orders in some grids, we made use of all-item orders to impute the electronics orders because they have a similar trend as shown in Fig.2.

2.2 Data Analysis

The number of fire incidents is impacted by several factors. In this section, we analyze two categories of factors: temporal effects and spatial factors.

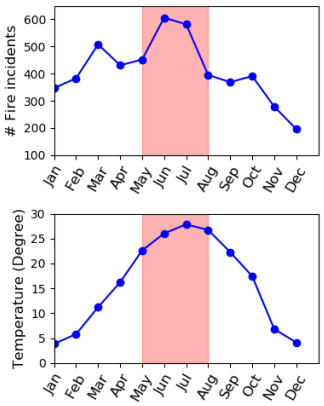
Temporal Effects. Temporal effects consist of two types: internal temporal effects (*e.g.*, historical fire incidents) and external temporal effects (*e.g.*, temperature, humidity and electronics orders).

Internal temporal effects. To determine whether historical fire risk will impact the current fire risk, we observe the correlation between the number of fire incidents in Nov. 2018 and the number of fire incidents in each month before Nov. 2018. The correlation is shown in Fig.3. The points with red edge mean significant correlation. We find some periodic impact on the number of fire incidents in Nov. 2018, with period of approximately 18 months. The two most recent months have the largest impact, while more distant months have smaller impacts on the number of fire incidents in Nov. 2018. These findings inform our determination of the sequence length of historical information.

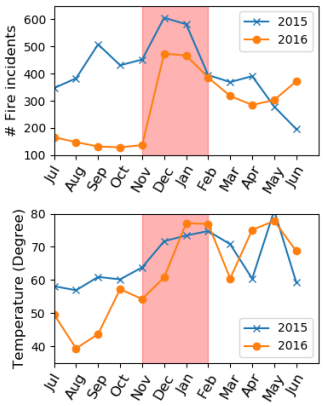
External temporal effects. We use three external temporal effects, which will influence the number of fire incidents, including temperature, humidity, and the number of electronic orders. The changes of these three factors and the number of fire incidents are shown in Fig.4. We analyze their impact as follows:

1) Immediate impact. From Fig.4(a), 4(b), 4(c) and 4(d), we learn that summer and winter each year are high-incidence seasons. Temperature and humidity seem to have an immediate impact. High temperature and high humidity appear to lead to more fires in summer as shown in the red areas in Fig.4(a) and Fig.4(b). This seems to contradict our common sense. The reason is that Henan province in China has a temperate continental climate. Higher temperatures in the summer are accompanied by higher humidity levels. On other hand, though the temperature is lower in winter, there is low humidity and an increased rise of fire startings in the dry air.

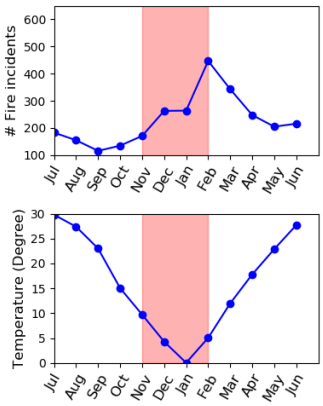
2) Delayed impact. Compared to the immediate impact of temperature and humidity on the number of fire incidents, the number of electronic orders has a delayed impact. From Fig.4(e), we find that the number of fire incidents increases after the electronic orders increasing. Because the peak of electricity use will come after the number of electronic products increasing. For example, numerous people buy air conditioners before summer.



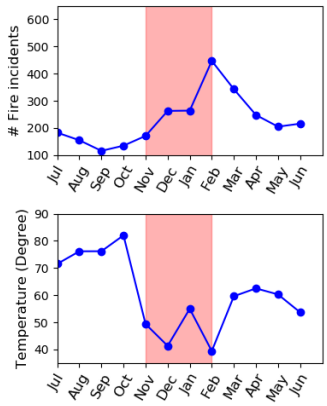
(a) Change in fire incidents and temperature in summer



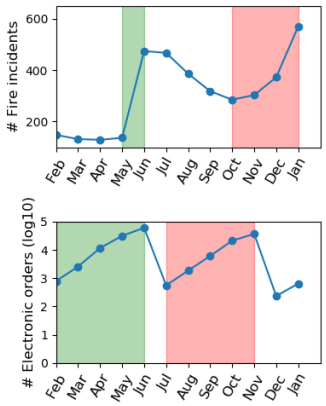
(b) Change in fire incidents and humidity in summer



(c) Change in fire incidents and temperature in winter



(d) Change in fire incidents and humidity in winter



(e) Change in fire incidents and electronic orders

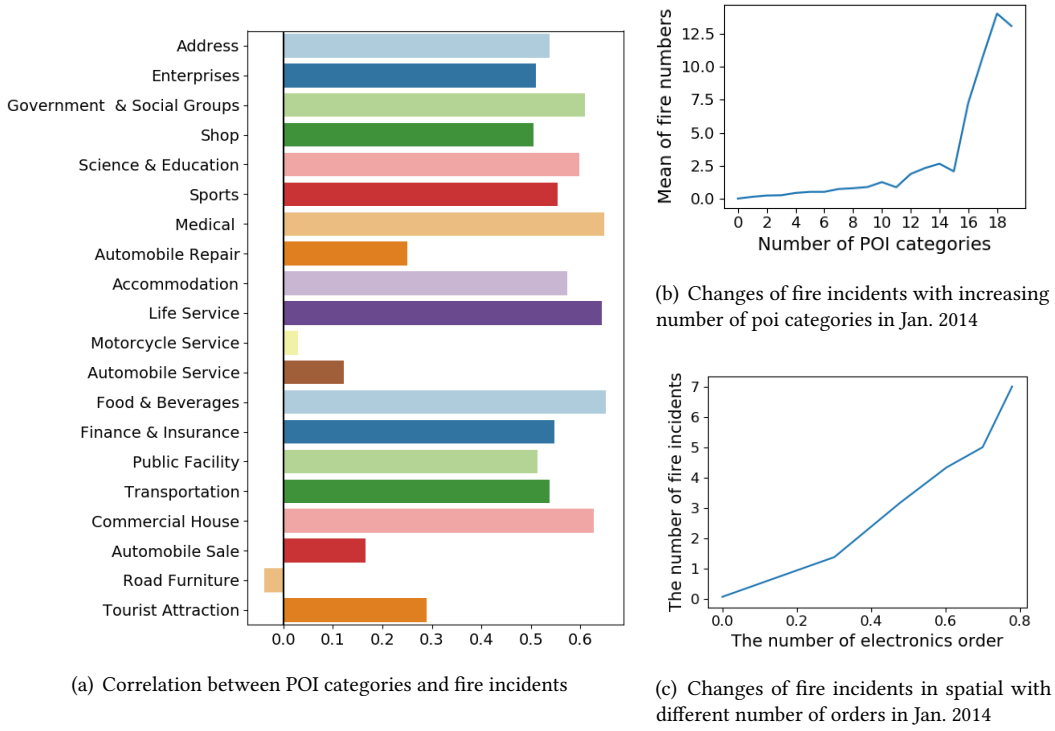
Fig. 4. Change in number of fire incidents and several temporal effects

But they use them in one or two months. From the red area and green area in Fig.4(e), we learn that the delay is about 3 months. It means that the increase in electronic orders precedes the increase in fire risk by three months. The increases in orders reflected in the red and green areas result in different patterns of increase in incidents of fire because the number of fire incidents is affected by several factors.

Spatial Factors Spatial factors include local spatial attributes and global dependencies.

Local spatial attributes. Local spatial attributes (*e.g.*, POI) will affect the fire risk at each timestamp and differentiate grids. For example, POIs are related to the crowd flow. And orders will reflect the usage of electricity consumption. As we all known, if there are more people or more electricity consumption in a grid, the possibility of fire risk will be high.

To analyze the relationship between the number of fire incidents and the poi diversity, we first counted the number of poi categories in each grid. Then we calculated the mean fire incidents of grids whose number of

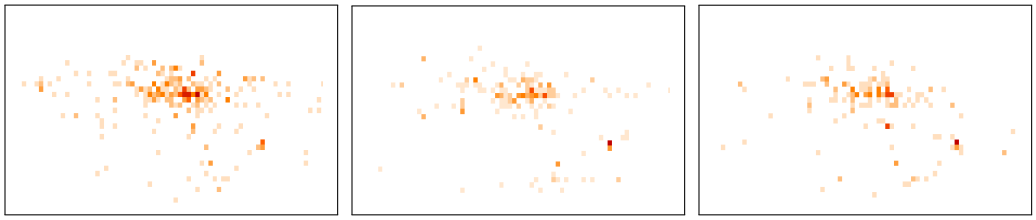


(a) Correlation between POI categories and fire incidents

(b) Changes of fire incidents with increasing number of poi categories in Jan. 2014

(c) Changes of fire incidents in spatial with different number of orders in Jan. 2014

Fig. 5. Relationship between poi categories and the number of fire incidents in spatial



(a) The distribution in March 2018 (b) The distribution in April 2018 (c) The distribution in May 2018

Fig. 6. The distribution of fire incidents varying from time

categories are same. Fig. 5(b) displays the changes of fire incidents number when the number of poi categories is increasing. Obviously, grids with increasing number of categories are easy to break out of fires. The reason may be that these grids draw more people, which increase the possibility of fire incidents. Meanwhile, we calculated the correlation between the number of each POI categories and the number of fire incidents as shown in Fig.5(a). The top five correlated POI categories are Food & Beverages, Medical, Life Service, Commercial House, Governments & Social groups. We also analyze the relationship between the number of orders and the number of fire incidents at one timestamp, using the same method with analyzing the number of POI categories. As shown in Fig. 5(c), we learn that the grids with more electronics orders are easier to break out of fires. The reason could be that more electronics will cause overuse of electric power, which results in fires.

Table 1. Mathematical Notations

Symbol	Description
$A = \{a_i\}_{i=1}^{n}$	a set of grids in the city
$N_t = \{N_{t,i}\}_{i=1}^{n}$	The fire records at t timestamp
$Y_t = \{y_{t,i}\}_{i=1}^{n}$	The fire risk of each grid at t timestamp
$X_t = \{x_{t,i}\}_{i=1}^{n}$	The features of each grid at time stamp
$M_t = \{m_t^{temp}, m_t^{hum}\}$	Meteorology features including temperature and humidity at t timestamp
$C_t = \{c_{t,i}^{l}, c_{t,i}^{e}\}_{i=1}^{n}$	The features of E-commerce for each grid at t timestamp including all item orders and electronic orders
$S = \{s_i\}_{i=1}^{n}$	Spatial features (*e.g.*, POI) for each grid

Global spatial dependencies. The distribution of fire incidents across the city demonstrates global dependencies that do not change significantly with time. Fig.6 shows the distribution of fire incidents in three adjacent months. Obviously, the regions in the center of the figures always have high fire risk and fires seldom break out in the regions in the margin of figures. We can leverage the fire risk difference between grids to discriminate the fire risk of grids across space.

From the analysis above, we find that the number of fire incidents is influenced by the combined action of immediate impact and delayed impact. Even when temperature is low and humidity is high, reducing the fire risk, fires will break out in some grids, notably those with highly correlated POIs or an increasing number of electronic orders. It is challenging to represent such complex factors jointly. We need to integrate the internal and external temporal effects to represent the changes of fire risks along a temporal dimension. Meanwhile, we should consider the spatial dependencies to learn the fire risk distribution in space at each timestamp.

3 PROBLEM STATEMENT

In this paper, we use features of last several timestamps to forecast the fire risk of current timestamp. We briefly define our problem as following:

Definition 1: We partition the region of a city into n grids (*e.g.*, 1km×1km) $A = \{a_1, a_2, ..., a_n\}$ based on latitude and longitude as shown in Fig.6.

Definition 2: At the t^{th} timestamp, the number of fire incidents in each grid is $\{N_{t,i}|N_{t,i} \geq 0, i = 1, 2, .., n\}$. Then we sort these grids in ascending order and normalize the index to (0,1). Let the normalized results $\{y_{t,i}|i = 1, ..n\}$ denote the fire risk assigned to the grids, which is defined as following:

$$\begin{cases} y_{t,i} < y_{t,j} & if N_{t,i} < N_{t,j} \\ y_{t,i} = y_{t,j} & if N_{t,i} = N_{t,j} \\ y_{t,i} = 0 & if N_{t,i} = 0 \end{cases} \tag{1}$$

where $y_{t,i} \propto N_{t,i}$. The effects of features are also proportional to the fire risk. Therefore, the analysis in Section 2.2 can also be applied in fire risk.

Definition 3: We extract spatial-temporal features $\{X_1, ..., X_t\}$. For each timestamp t_k, $X_{t_k} = \{M_{t_k}, C_{t_k}, S\}$, where M_{t_k}, C_{t_k}, S is defined in Table 1. To extract the features, we normalize the value of Meteorology (*e.g.*, temperature, humidity), E-commerce data(*e.g.*, the number of electronic orders and all item orders) and spatial features (*e.g.*, the number of POI categories) to (0, 1) respectively.

Given a Δt period of historical fire risks $\{Y_{t-\Delta t}, ..., Y_{t-1}\}$ and spatial-temporal features $\{X_{t-\Delta t}, ..., X_{t-1}\}$, we forecast the t^{th} timestamp fire risk Y_t.

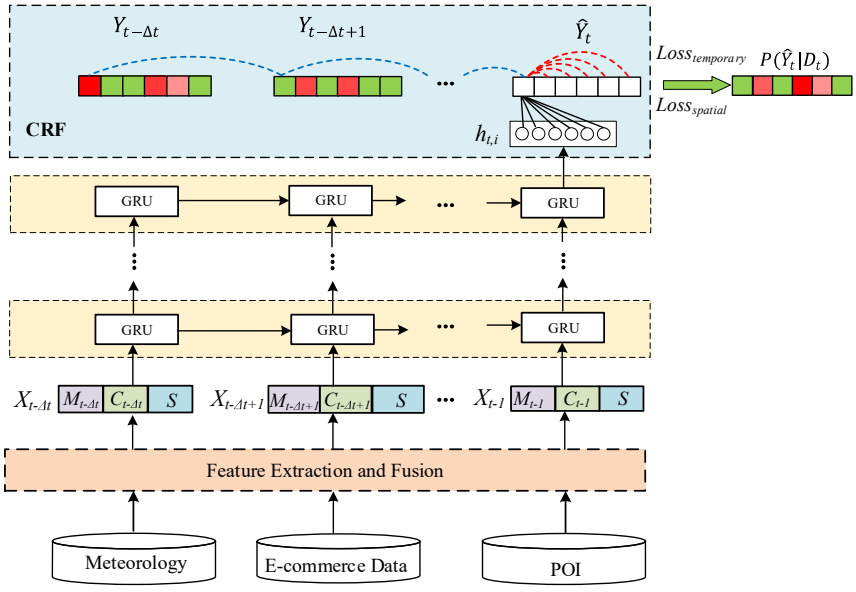


Fig. 7. Architecture of NeuroFire.

4 FIRE RISK FORECASTING MODEL

Because of the difficulties of forecasting on such a sparsity datasets, we propose a machine learning model named NeuroFire to integrate the temporal and spatial (Shown in Fig.7), which consists of two-step: temporal fire classification and spatial fire risk forecasting .

4.1 Temporal Fire Classification

From the data analysis, we learn that the current fire risk is affected by two kinds of effects from temporal view: internal temporal effects and internal temporal effects (as mentioned in Section 2.2). The internal temporal effects shows the correlation with historical fire risks. And the internal temporal effects influence the fire risk by immediate impact and delay impact. To forecast the current fire risk, we use GRU-CRF [15] to combine these two kinds of factors. Specifically, we first use GRU to learn temporal representation of internal temporal effects and use CRF to combine the internal internal temporal effects and temporal representation.

Recently, CRF achieves success in temporal sequence inference in urban, such as crime forecasting [37] and air pollution forecasting [43]. Different from linear HMM depending on the last state, CRF depends on the sequence before, which avoid label bias. A linear Conditional Random Field (CRF) combines these two kinds of factors by building an undirected graph. Additionally, CRF has the capacity of avoiding label bias, which is helpful for our problem. Because the label number of no fires is much greater than the label number of fires. If label bias happens, the most fires labels will be forecasted as no fire. Due to sparsity of fires labels, we transform regression problem to binary classification. It means that we will forecast whether there are fires in grid i, which is denoted by $I(\cdot)$, where $I(y) = 1, if\ y > 0$ and $I(y) = 0,\ otherwise$. The probability of $I(y_{t,i})$, which represents the fire risk of grid i, given urban features $D_{t,i} = \{x_{t-\Delta t,i}, ..., x_{t-1,i}\}$ and historical fire risk $\{y_{t-\Delta t,i}, ..., y_{t-1,i}\}$ is defined as (2).

$$P(I(y_{t,i})|D_{t,i}) = \frac{1}{Z(D_{t,i})} exp\left(\sum_p \alpha_p f_p(I(y_{t,i}), D_{t,i}) + \sum_{k=t-\Delta t}^{t-1} \sum_p \beta_p g_p(I(y_{k+1,i}), I(y_{k,i}), D_{t,i})\right) \quad (2)$$

where f_p and g_p are linear functions. $Z(D_{t,i})$ is a normalization factor. All grids share the parameters.

However, some of urban features cannot directly effect the current result because of the delay. Therefore, We leverage Gate Recurrent Unit (GRU) to address this problem, which has been employed in a variety of sequential learning works. The GRU can explore long-range temporal dependencies by maintaining or forgetting a memory based on historical information. Compared to LSTM, it is lightweight and efficient. This idea fits our problem. For example, some features like the number of orders have some delays on affecting fire risk. On the other hand, fire risk will increase in the short term when temperature and humidity are high. The hidden output $h_{t,i} \in \mathbf{R}^{H_1}$ is temporal representation of $D_{t,i}$, which is calculated by (3) given $D_{t,i}$:

$$\begin{aligned} h_{t-\Delta t+1,i} &= GRU(x_{t-\Delta t,i}) \\ &\cdots \\ h_{t,i} &= GRU(x_{t-1,i}, h_{t-1,i}) \end{aligned} \tag{3}$$

After computing the hidden output, we need to calculate the emission probability of each dimension as the input of CRF. To fit the input of CRF, we let $h_{t,i}$ go through two full connection layers as following:

$$\begin{aligned} o^1_{t,i} &= w_{fc1} \cdot h_{t,i} + b_{fc1} \\ o^2_{t,i} &= w_{fc2} \cdot o^1_{t,i} + b_{fc2} \end{aligned} \tag{4}$$

where $w_{fc1} \in \mathbf{R}^{H_2 \times H_1}$, $b_{fc1} \in \mathbf{R}^{H_2}$, $o^1_{t,i} \in \mathbf{R}^{H_2}$, $w_{fc2} \in \mathbf{R}^{2 \times H_2}$, $b_{fc2} \in \mathbf{R}^2$, $o^2_{t,i} \in \mathbf{R}^2$. Each full connection is used to reduce the dimension of $h_{t,i}$, which means $H2 < H1$.

Through the GRU and full connection layers, we transform the $X_{t,i}$ into $o^2_{t,i}$, which including a Δt period of information. Combining GRU and CRF, we re-formulate (2) as following:

$$P(I(y_{t,i})|D_{t,i}) = \frac{1}{Z(D_{t,i})} exp\left(\sum_p \alpha_p f_p(I(y_{t,i}), o^2_{t,i}) + \sum_{k=t-\Delta t}^{t-1} \sum_p \beta_p g_p(I(y_{k+1,i}), I(y_{k,i}), o^2_{t,i})\right) \tag{5}$$

where $P(I(y_{t,i}) = 1|D_{t,i})$ is the probability of that there are fires, which means fire risk.

The loss function in temporal fire forecasting is motivated by MLE as following:

$$Loss_{temporal} = \arg\max P(I(y_{t,i})|D_{t,i}) \tag{6}$$

4.2 Spatial Fire Risk Forecasting

Though temporal fire forecasting calculates the fire risk (probability) of fires for each grid, the fire risks of grids are independent. The probability of fires cannot provide evidence of discriminating fire risk of grids at one timestamp. To learn spatial dependencies, we need to consider the fire risk differences among grids.

This problem is similar to the ranking problem in the recommendation system. In order to recommend products for users, Rendle *et al.* [27, 28] proposed S-BPR method to discriminate products by partitioning products into two sets: the products that are preferred by users and other products.

However, there is a difference from the problem and the product recommendation that we need to forecast fire risks of the grids at the t^{th} timestamp. Inspired by S-BPR [27, 28], we transform the fire risk forecasting problem into a pair-wised comparison in spatial. For each grid i, other grids can be classified into two classes. The one class contains the grids that the fire risks are greater than the grid i. The other one contains remaining grids.

For example, as shown in Fig.8, Y_t are fire risks labels for grids at the t^{th} timestamp. We transform the metric for Y_t into a $n \times n$ metric r_t, which is a label for dependencies between grids, where $r_t(i,j) = 1, if\ y_{t,i} > y_{t,j}$ and $r_t(i,j) = 0,\ otherwise$. Similar to the operation for fire risks labels, the dependencies of forecasting scores is $\hat{r}_t(i,j) = 1, if\ p_{t,i} > p_{t,j}$, where $p_{t,i} = P(\hat{y}_{t,i} = 1|x_{t-1})$. $\hat{y}_{t,i} \in \{0,1\}$ represents the forecasting result of whether there are fires at the t^{th} timestamp. These two dependencies metrics are calculated by (7).

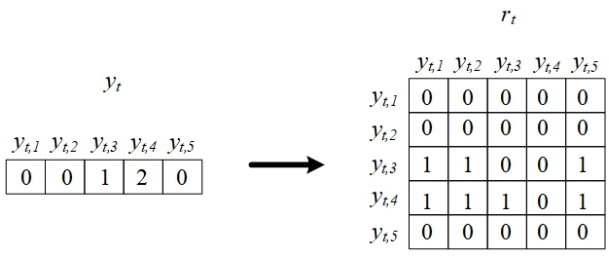


Fig. 8. Illustration of spatial dependencies

$$r_t = I\left(\begin{bmatrix} Y_t^T & -\mathbf{1} \end{bmatrix}\begin{bmatrix} \mathbf{1} \\ Y_t \end{bmatrix}\right) \qquad \hat{r}_t = \sigma\left(\begin{bmatrix} p^T & -\mathbf{1} \end{bmatrix}\begin{bmatrix} \mathbf{1} \\ p \end{bmatrix}\right) \tag{7}$$

We adopt cross entropy as the objective function in spatial fire risk forecasting in (8). This step is used to adjust $p_{t,i}$, which makes the fire risks rank sorted by $p_{t,i}$ close to the fire risks rank sorted by fire risk in real world.

$$Loss_{spatial} = \arg\max \sum_k \sum_l r_t(k,l)\log(\hat{r}_t(k,l)) + (1 - r_t(k,l))log(1 - \hat{r}_t(k,l)) \tag{8}$$

Finally, we combine the $Loss_{spatial}$ and $Loss_{temporal}$ to forecast fire risks of grids at the t^{th} timestamp. The actual objective function we optimize is $Loss_{st} = Loss_{temporal} + Loss_{spatial}$

5 EXPERIMENT

5.1 Experiment Setup

We conduct experiments on dataset including 59-month fire incidents and urban data as described in Section 2.1. We divided Zhengzhou into 1km×1km grids (5472 grids). Over 96% grids have no fires in each month. For cross validation, we use a sliding window to divide the dataset into train sets, validation sets and test sets along the time sequence. The time length of the window are 49 months, including 42 months for train set, 6 months for validation set and 1 month for test set. For example, if we forecast the fire risk in November 2018 (test set), the train set is from November 2014 to April 2018 and the validation set is from May 2018 to October 2018. In experiments, we forecast the fire risk in 11 different months (January 2018 to November 2018). Due to small dataset, we use only one GRU layer. When train the model, we select Adam [18] optimizer. Meanwhile, we set the hidden size of GRU unit as 32 and set the learning rate as 0.001.

5.1.1 Metrics. We use three metrics to evaluate our model: F1-score, AUC and MAE as following:

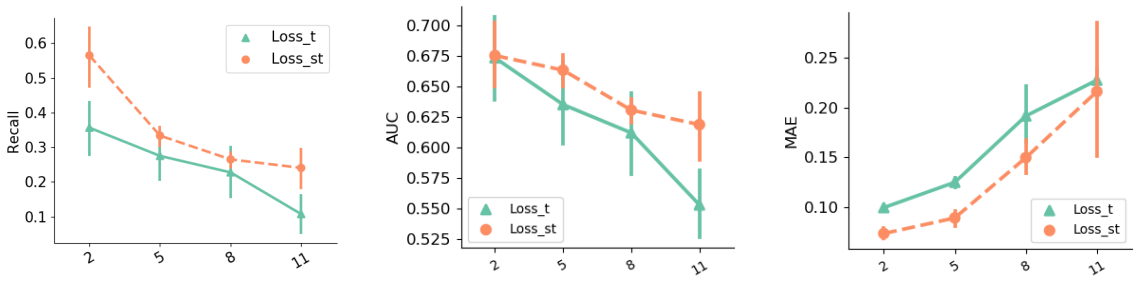
$$F1 - score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{9}$$

$$precision = \frac{\sum_i^n I_m(p_{t,i} > P(\hat{y}_{t,i} = 0|x_{t-1}) \wedge y_{t,i} > 0)}{\sum_i^n I_m(p_{t,i} > P(\hat{y}_{t,i} = 0|x_{t-1}))}, \quad recall = \frac{\sum_i^n I_m(p_{t,i} > P(\hat{y}_{t,i} = 0|x_{t-1}) \wedge y_{t,i} > 0)}{\sum_i^n I_m(y_{t,i} > 0)} \tag{10}$$

where $I_m(\cdot)$ is indicative function. $I_m(s) = 1$, if s is True and $I_m(s) = 0, otherwise$.

$$AUC = \frac{1}{|E(t)|} \sum_{i,j \in E(t)} I_m(\hat{y}_{t,i} > \hat{y}_{t,j}), \quad E(t) = \{(i,j)|I(y_{t,i}) > I(y_{t,j})\} \tag{11}$$

$$MAE = \frac{1}{n}\sum_i^n |y_{t,i} - p_{t,i}| \tag{12}$$



(a) Recall (b) AUC (c) MAE

Fig. 9. Impact of the length of period under different loss function. (Loss_t represents $Loss_{temporal}$.)

5.1.2 Baseline models. To evaluate NeuroFire, we select 9 baselines models. All baselines as following were run 10 times:

- **CRF:** Conditional Random Field model leverages transition probabilities between labels in sequence and emission probabilities from features to labels in order to calculate the score of labels.
- **LASSO:** Least absolute shrinkage and selection operator model is a regression model which fits the problem of forecasting fire risks of grids.
- **SVM [9]:** Support vector machine is always used in classification, which performed best in Firebird [22], which is a work about fire risk prediction. We only use features of one timestamp for SVM.
- **SVM-P:** To find whether long-term features improve the performance of SVM, we concatenate 2 months of features for SVM named as SVM-P.
- **LR:** Logistic regression model is always used for classification.
- **LR-P:** Similar to SVM-P, we concatenate 2 months of features as input of LR.
- **GRU [6] (GRU-LR):** We implement GRU to fit our two-step method. The hidden output is contributed to the input of LR, which transforms LR into a sequential learning model by using GRU to learn temporal delays.
- **GRU-CRF [15]:** It has the similar architecture as our model, but ignore spatial dependencies.
- **DeepST [40]:** It divides the time series into three types of temporal components: closeness, period and trend, extracts spatial features at each timestamp of each components by CNN.

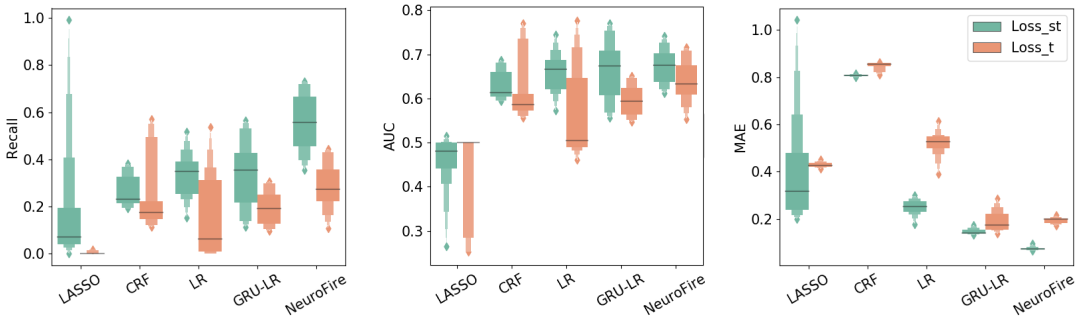
5.2 Performance For Period of features

The fire risks at the t^{th} timestamp are contributed by historical fire risks sequence and urban feature in a Δt period. The length of the period determines the richness of historical information. Because the longer period resulting in richer or more useless information is unknown We evaluate the length of the period to find the appropriate Δt.

We select 4 different values (2, 5, 8, 11) for Δt, and evaluate them under two different loss function ($Loss_{temporal}$ and $Loss_{st}$). The results are shown in Fig.9. With increasing length of period, the performance of our model under two loss functions both go down. The results infer that historical sequence and urban features contain some useless information even some noise when the length of period is long. When $\Delta t = 2$, our model under two different loss functions both reach the best result. These results show that the fire risks for current month are significantly related to the two most recent months as mentioned in Section 2.2.

Table 2. Comparison with baseline models

Model	**Recall**	**F1-score**	**AUC**	**MAE**
LASSO	0.202±0.306	0.025±0.014	0.454±0.075	0.411±0.265
CRF	0.263±0.075	0.349±0.071	0.629±0.037	0.807±0.005
LR	0.333±0.59	0.324±0.059	0.659±0.052	0.249± 0.036
LR-P	0.179 ±0.188	0.121±0.058	0.573±0.072	0.162±0.019
SVM	0.540±0.273	0.087±0.038	0.56±0.096	0.498±0.224
SVM-P	0.308±0.066	0.227±0.085	0.629±0.031	0.267 ±0.097
GRU-LR	0.342±0.101	0.365±0.102	0.666±0.076	0.146± 0.0139
GRU-CRF	0.284±0.111	0.387±0.071	0.673±0.06	0.100±0.003
DeepST	0.532±0.070	0.310±0.080	0.740±0.037	0.193±0.092
NeuroFire	**0.558±0.134**	**0.400±0.067**	**0.763±0.045**	**0.094±0.01**



(a) Recall (b) AUC (c) MAE

Fig. 10. Comparison with models using two loss functions. (Loss_t represents $Loss_{temporal}$.)

5.3 Performance of Our Model

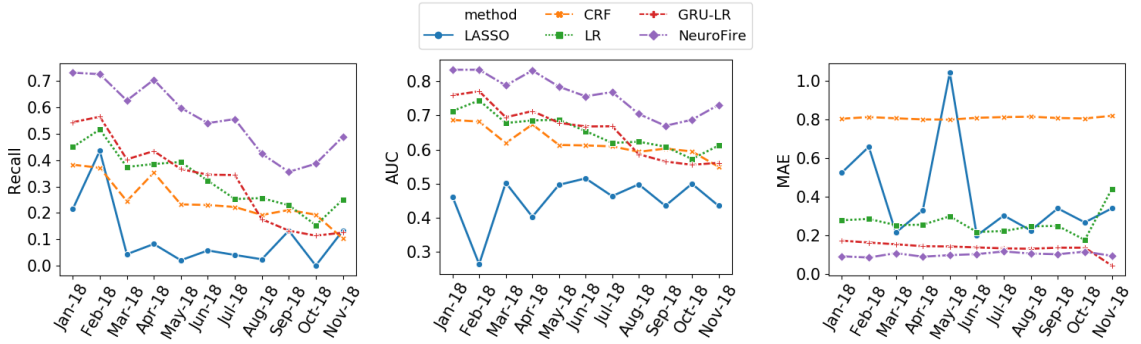
5.3.1 Comparison with different baselines. Table 2 shows the performances of our model (NeuroFire) and baseline models. The results in table are mean and standard deviation after running 10 times. Baseline models except DeepST are trained by adding $Loss_{spatial}$ because they don't consider spatial dependencies. Our model produces the best performance. From the results, DeepST doesn't perform good on the situation that spatial distribution is sparse, which makes difficult to train CNN. Moreover, DeepST is a much more complicated model that needs bigger datasets to train. GRU-LR and GRU-CRF perform better than LASSO, CRF, LR and SVM. Because GRU-LR and GRU-CRF consider temporal features. Though LR-P and SVM-P involve more than one timestamp of features, they still perform worse than GRU-LR and GRU-CRF. Because GRU not only keeps long-term information but also discards some useless information. In addition, GRU-CRF perform better than GRU-LR because of considering historical sequence of fire risks. LASSO performs worst because the labels of fire risk are very sparsity. It makes difficult to make regression. We check the results of forecasting using LASSO. Then we find that almost every grid is forecasted as the same low risk even 0. Compared to LASSO, LR performs better in classification but cannot do well in spatial distinction either. The results show that historical sequence and temporal delays are both important to forecast the fire risks.

5.3.2 Performance for involving $Loss_{spatial}$. To evaluate the impact of $Loss_{st}$, we train our model (NeuroFire) and 4 baseline models (*e.g.*, LASSO, CRF, LR, GRU-LR) by using two loss functions. $Loss_{temporal}$ for each baseline is their own loss functions. $Loss_{st}$ for each baselines means that adding $Loss_{spatial}$ to $Loss_{temporal}$. In particular,

our model only using $Loss_{temporal}$ is GRU-CRF in baseline models. The results are shown in Fig.10. Compared to models using $Loss_{temporal}$, models using $Loss_{st}$ perform better in Recall, AUC and MAE. It infers that $Loss_{st}$ improves not only the results of finding top risk but also the results of forecasting. In summary, models using $Loss_{st}$ outperform than models using $Loss_{temporal}$.

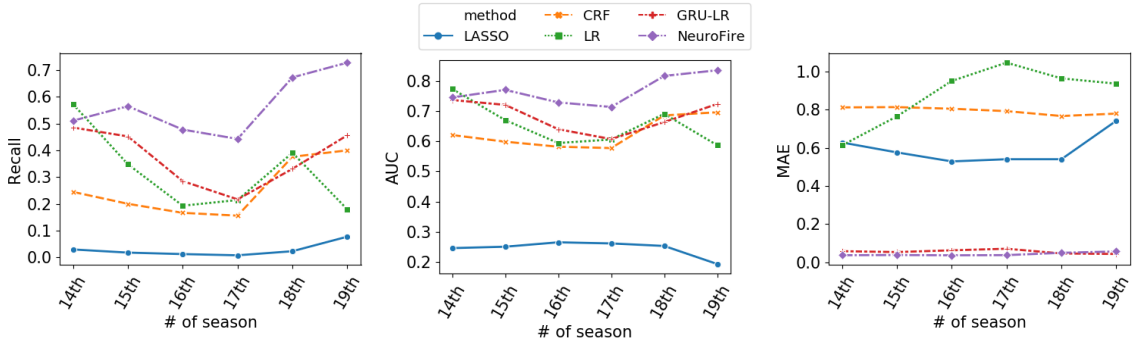
5.3.3 Performance on different situations. To evaluate the stability of our model, we conduct experiments on different time scales and sizes of grids.

Performance on different time scales. The results for each month is shown in Fig.11. We conduct experiments on seasonal data (three months data). 59 months are divided into 19 seasons. We also used a sliding window to select train sets (12 seasons), validation set (1 season) and test set (1 season). We forecast the fire risk from the 14th season to 19th season as shown in Fig.12. Compared to forecasting monthly fires, our model performs better on forecasting seasonal fires. Recall for seasonal forecasting is over 40%. It is because that the features of seasons are more distinct than the features of months. From the results, we find that our model outperforms on both month-level and season-level. Because we not only consider the long-term features but also involve the historical fire risk sequence, which is useful to avoid label bias in such a sparse dataset.



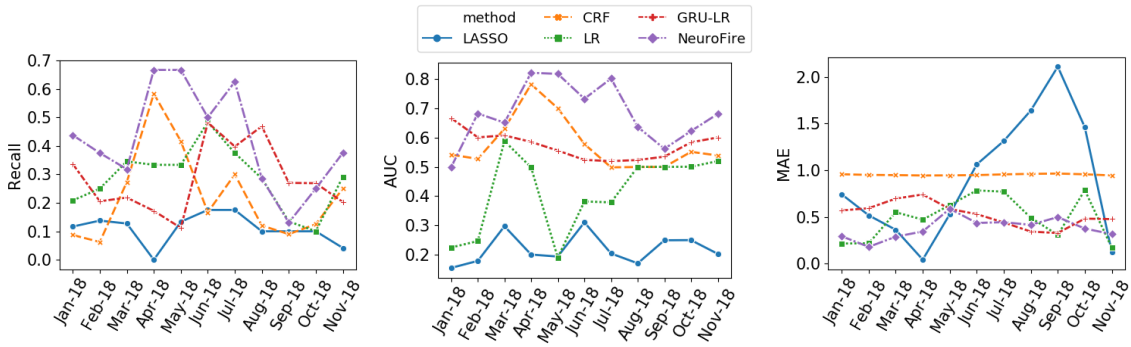
(a) Recall (b) AUC (c) MAE

Fig. 11. Performance for each month. The number of grids with fires from Jan.2018 to Nov.2018 is 149, 124, 139, 122, 112, 87, 99, 125, 166, 176, 256.



(a) Recall (b) AUC (c) MAE

Fig. 12. Performance for each season. The number of grids with fires from 14th season to 19th season is 274, 230, 337, 441, 266, 195.



(a) Recall (b) AUC (c) MAE

Fig. 13. Performance for each month when grids are small.

Table 3. Performance related to features

Features	Recall	F1	AUC	MAE
w/o m_t^{temp}	0.288±0.072	0.368±0.064	0.641±0.035	0.114±0.023
w/o m_t^{hum}	0.334±0.127	0.372±0.078	0.662±0.061	0.094±0.011
w/o c_t^l	0.376±0.123	0.391±0.056	0.682±0.058	0.074±0.013
w/o c_t^e	0.368±0.117	0.389±0.062	0.678±0.056	0.111±0.016
w/o S	0.362±0.141	0.373±0.068	0.675±0.067	0.06±0.006
$m_t^{temp} + m_t^{hum} + c_t^l + c_t^e + S$	**0.558±0.134**	**0.400±0.067**	**0.763±0.045**	**0.094±0.01**

Performance on different sizes of grids. To evaluate stability of our model, we also conduct experiments on smaller grids which are 0.5km×0.5km (21080 grids in total). The sparsity for each month is over 98.2% As shown in Fig.13, the performances of each model looks worse than results based on 1km×1km grids due to the higher sparsity. Our model still performs better than other models.

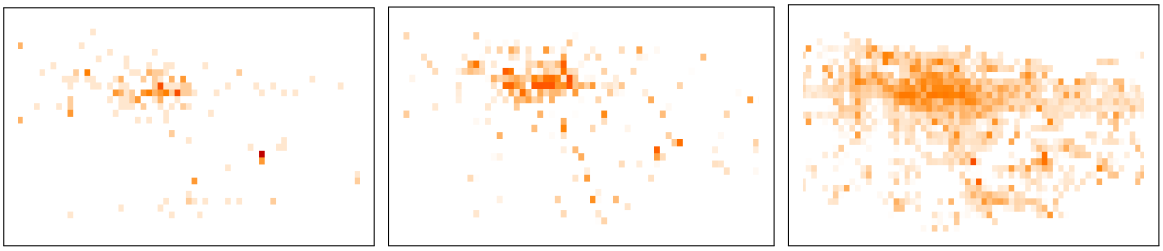
5.4 Evaluation on Features

To find the feature which plays an important role, we study the performance of model by different combinations of features as shown in Table 3. We learn that the performance is lowest when taking out one of the features. Recall and F1 is significantly improved when involved temperature (m_t^{temp}). Therefore, we believe that temperature is easy to influence the fire risk. Humidity (m_t^{hum}) also plays an important role in determining fire risk. Recall decreases 22.4% when taking humidity (m_t^{hum}) away.

5.5 Visualization of Results

Visualization for performances of Loss_{st}. We demonstrate the result of our model when using $Loss_{temporal}$ and $Loss_{st}$ as shown in Fig.14. Compared to using $Loss_{st}$, our model using $Loss_{temporal}$ forecasts some wrong risks for some grids, which will not break out of fires. In addition, the rank result when using $Loss_{temporal}$ is not accurate. Many grids cannot be discriminated. In summary, the visual result supports the conclusion in Experiment 5.3.2 that considering spatial dependencies improves the forecasting.

Visualization for results of models. The visualization for forecasting results of one month is shown in Fig.17. The deeper color in a grid, the higher risk is forecasted. The results demonstrate that our model perform better in forecasting. GRU-LR and LR outperform than CRF, LASSO and SVM. Though the visualization result of



(a) Groundtruth in Jan. 2018 (b) The result when using $Loss_{st}$ (c) The result when using $Loss_{temporal}$

Fig. 14. Demonstration the results of NeuroFire when using $Loss_{temporal}$ and $Loss_{st}$.

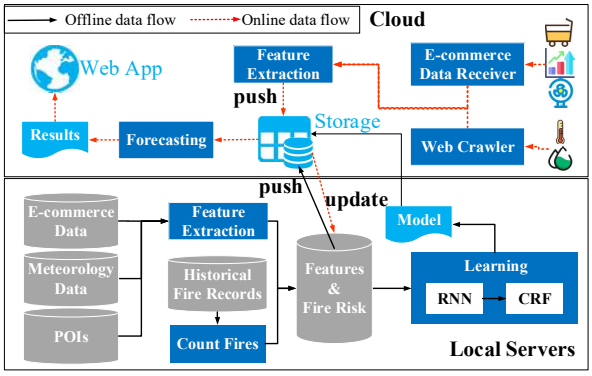
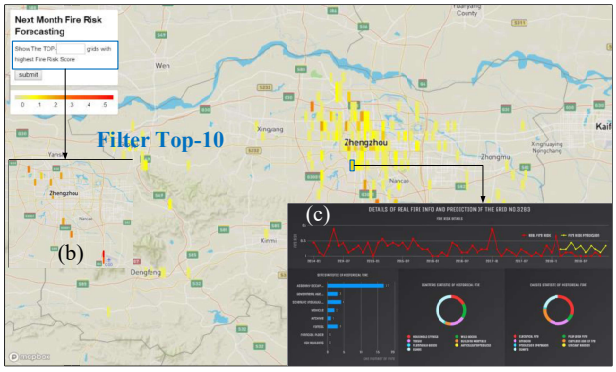


Fig. 15. System Framework of CityGuard



(a)

Fig. 16. Interface of CityGuard

DeepST seems discriminating in spatial, it performs worse in forecasting. In summary, NeuroFire outperforms other baselines in forecasting.

6 SYSTEM ARCHITECTURE

In order to proof government's fire prevention and monitoring work, we develop a monitoring and forecasting system to intuitively represent historical fire information and predictions of our model. The CityGuard's system framework consists of two major parts as shown in Fig.15: *local servers* and the *Cloud*.

Local Servers Local servers mainly handle the offline time-consuming training (learning), including three tasks:

- Extract and construct features from urban database: we first extract a period of temporal and spatial features for each grid then fit features to a feature construct module to get well-format features and store them in local.
- Count historical fire records of each grid: we first collect the historical fire records from government and apply a calculating module to get fire risk data for each grid, then store them in local.
- Training model: we use above fire risk data and features to train a forecasting model via our model, and then upload the learned model to the *Cloud*.

The Cloud The *Cloud* continuously crawls meteorological data and monthly receives E-commerce data. After extracting features from data, the *Cloud* stores them. With the learned model and new features data, the *Cloud*

computes the fire risk for each and every grid of a city. The web service in the *Cloud* share a App Service to visualize predicted fire risk.

Interface Fig.16(a) presents the website of our Fire Risk Forecasting System (CityGuard)[6]. The color of each grid shows according to the degree of its fire risk, *e.g.*, "yellow" represents "low" fire risk and "red" represents "high" fire risk. The top-left corner of the website shows the input box with submit button in which the user can type a number. This number represents the number of grids with the highest risk that user wants to focus on in the whole grids list. For example, as shown in Fig.16(b), a heatmap with only top-10 highest fire risk grids has shown if user inputs the number 10. A user can also click any grid with colors on the website to see the detail information. Fig.16(c) presents charts of historical and forecasting fire detail information: the variation of fire risk over time (at the top), the historical statistic of types of fire sites (at bottom-left), igniters (at bottom-center) and causes (at bottom-right).

7 RELATED WORK

7.1 Forecasting in Urban Computing

Recently, urban data that is generated in cities (*e.g.*, POIs, E-commerce data, APP data[24]) is widely used to forecast urban states, such as air pollution forecasting [10, 20, 38, 43], site recommendation [4, 13, 21], mobility forecasting [11, 12, 17, 35, 39–41], and crime prediction [5, 14, 42]. The various kinds of urban data show the potential of assisting us with urban planning. To explore the laws of cities, many works propose temporal models to characterize the dynamic changes of city. For example, Devarakonda *et al.* [10] leverage a sensing device to measure carbon monoxide in order to predict pollution using linear regression. Xu *et al.* [34] represent the temporal mode of each mobile user by partitioning their trajectories to explore the different life style of users. Cao *et al.* [2] identify three main revisitation patterns based on check-in data. And they explore living pattern by embedding trajectories [3]. Fang *et al.*[12] extract some features from multiple cellphone network then fuse the population estimator for each network. With widely use of neural network, many works study on spatial-temporal forecasting by combining Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), which can reduce the effort of feature selection. Qin *et al.* [25] leveraged CNN to extract the spatial features then learned temporal dependencies using LSTM. Shi *et al.* [32] extended the fully connected LSTM to have convolutional structures in both the input-to-state and state-to-state transitions. The methods perform well in urban computing. However, they use spatial features for each grid to forecast but do not consider about the global spatial dependencies, which is used to determine the priority in rank list.

7.2 Urban Fire Risk Forecasting

Existing fire forecasting methods focus on building Risk-based Fire Inspection System. For early works, they prioritized the inspection objects without temporal impact. The New York Mayor's Office of Data Analytics with the Fire Department of New York (FDNY) firstly developed a Risk-Based Inspection System (RBIS). They constructed a data-driven fire forecasting model to identify and prioritize these property inspections according to a set of buildings and behaviors information [8]. Lately, the "Firebird" was developed to identify and prioritize commercial property inspections by involving more features of building-level [22, 23]. They utilized SVM and random forest based on historical fire incidents data and commercial data to prioritize the commercial properties. In fact, these existing works forecast the fire risk of several buildings based on but not temporal sequences of features, which ignores some changes along time. To address this challenge, Singh Walia *et al.*[29] updated the features weekly from various open sources. But they didn't consider the dependencies of temporal sequence.

[6]The access of website: http://101.124.0.58:5000/. The interface of our system uses open source tools, such as Mapbox (www.mapbox.com) and HighCharts (www.highcharts.com)

Our work is different with existing works from two aspects. Firstly, the works above ignored the dependencies between the temporal impact. We aim to characterize temporal dependencies from temporal features by RNN. Secondly, different from existing works, we consider the spatial dependencies which are beneficial to prioritize the area with top risk. Thirdly, existing works cost much time on feature engineering, for example, they extract and analyze over 200 features. Such demanding pre-process is difficult to apply in city-scale. Compared to these methods, our method can reduce such effort by using neural networks.

7.3 CRF Models in Spatial-temporal Forecasting

Conditional Random Field (CRF) [19] is mainly used in Natural Language Processing (NLP) [1, 7, 26] and computer vision [33]. In last decades, CRF is widely used in urban computing due to the performance of spatial-temporal forecasting. SGCRF [31] pruned redundant connection in fully connect CRF to forecast time-series energy. Wang *et al.* [30] improved SGCRF by integrating RNN to learn temporal features. Yi *et al.* [36, 37] using a tree structure clustering based on CRF to cluster high similar regions into groups, which was used to forecast the crime incidence. Zheng *et al.* [43] leveraged CRF to model the temporal sequence as a part of spatial-temporal model. However, it mainly deals with the temporal dependencies on the sequence of targets but not pays attention to the temporal dependencies on the sequence of features, which is necessary in fire forecasting. For example, the continuous increasing of electronic orders will result an increasing of fire risk.

8 CONCLUSION AND FUTURE WORK

In this paper, we first analyze the temporal (*e.g.*, historical fire incidents, temperature and humidity) and spatial factors (*e.g.*, POIs) which impact on the fire risk. To forecast fire risk, we leverage GRU-CRF model to learn temporal dependencies and design a spatial-temporal loss function to learn spatial dependencies at a timestamp. We evaluate our model on a real-world fire incidents and compare with 9 baseline models. Meanwhile, we depict the results of each models. The results show that our model outperforms 9 baseline models in fire risk forecasting.

From some existing works about fire risk analysis, we learn that many individual and group behaviors show correlation the fire risk [16]. In the future, we will involve more urban data (*e.g.*, consumption, crowd flows) to discover more factors directed to fire risk and improve accuracy of our model.

ACKNOWLEDGMENTS

This work was partially supported by the National Key R&D Program of China(2017YFB1001800), and the National Natural Science Foundation of China (No. 61772428,61725205,61902320,61972319). Finally, thank Chentian Jin for polishing paper.

REFERENCES

[1] Anurag Arnab, Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Måns Larsson, Alexander Kirillov, Bogdan Savchynskyy, Carsten Rother, Fredrik Kahl, and Philip H. S. Torr. 2018. Conditional Random Fields Meet Deep Neural Networks for Semantic Segmentation: Combining Probabilistic Graphical Models with Deep Learning for Structured Prediction. *IEEE Signal Process. Mag.* 35, 1 (2018), 37–52. https://doi.org/10.1109/MSP.2017.2762355

[2] Hancheng Cao, Zhilong Chen, Fengli Xu, Yong Li, and Vassilis Kostakos. 2018. Revisitation in Urban Space vs. Online: A Comparison Across POIs, Websites, and Smartphone Apps. *IMWUT '18* 2, 4, Article 156 (Dec. 2018), 24 pages. https://doi.org/10.1145/3287034

[3] Hancheng Cao, Fengli Xu, Jagan Sankaranarayanan, Yong Li, and Hanan Samet. 2019. Habit2vec: Trajectory Semantic Embedding for Living Pattern Recognition in Population. *IEEE TMC* (2019). https://doi.org/10.1109/TMC.2019.2902403

[4] Longbiao Chen, Daqing Zhang, Leye Wang, Dingqi Yang, Xiaojuan Ma, Shijian Li, Zhaohui Wu, Gang Pan, Thi Mai Trang Nguyen, and Jérémie Jakubowicz. 2016. Dynamic cluster-based over-demand prediction in bike sharing systems. In *UbiComp '16*. 841–852. https://doi.org/10.1145/2971648.2971652

[5] Xinyu Chen, Youngwoon Cho, and Suk Young Jang. 2015. Crime prediction using Twitter sentiment and weather. In *2015 Systems and Information Engineering Design Symposium*. 63–68. https://doi.org/10.1109/SIEDS.2015.7117012

[6] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).

[7] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel P. Kuksa. 2011. Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research* 12 (2011), 2493–2537. http://dl.acm.org/citation.cfm?id=2078186

[8] Eddie Copeland. 2015. BIG DATA IN THE BIG APPLE. In *https://policyexchange.org.uk/wp-content/uploads/2016/09/big-data-in-the-big-apple.pdf*.

[9] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks". *Machine Learning* 20, 3 (1995), 273–297. https://doi.org/10.1007/BF00994018

[10] Srinivas Devarakonda, Parveen Sevusu, Hongzhang Liu, Ruilin Liu, Liviu Iftode, and Badri Nath. 2013. Real-time Air Quality Monitoring Through Mobile Sensing in Metropolitan Areas. In *UbiComp '13*. ACM, New York, NY, USA, Article 15, 8 pages. https://doi.org/10.1145/2505821.2505834

[11] Zipei Fan, Xuan Song, Tianqi Xia, Renhe Jiang, Ryosuke Shibasaki, and Ritsu Sakuramachi. 2018. Online Deep Ensemble Learning for Predicting Citywide Human Mobility. *IMWUT '18* 2, 3, Article 105 (Sept. 2018), 21 pages. https://doi.org/10.1145/3264915

[12] Zhihan Fang, Fan Zhang, Ling Yin, and Desheng Zhang. 2018. MultiCell: Urban Population Modeling Based on Multiple Cellphone Networks. *IMWUT'18* 2, 3, Article 106 (Sept. 2018), 25 pages. https://doi.org/10.1145/3264916

[13] Bin Guo, Jing Li, Vincent W. Zheng, Zhu Wang, and Zhiwen Yu. 2018. CityTransfer: Transferring Inter- and Intra-City Knowledge for Chain Store Site Recommendation Based on Multi-Source Urban Data. *IMWUT '18* 1, 4, Article 135 (Jan. 2018), 23 pages. https://doi.org/10.1145/3161411

[14] Chao Huang, Junbo Zhang, Yu Zheng, and Nitesh V. Chawla. 2018. DeepCrime: Attentive Hierarchical Recurrent Networks for Crime Prediction. In *CIKM'18t (CIKM '18)*. ACM, New York, NY, USA, 1423–1432. https://doi.org/10.1145/3269206.3271793

[15] Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional LSTM-CRF Models for Sequence Tagging. *CoRR* abs/1508.01991 (2015), 1–10. arXiv:1508.01991 http://arxiv.org/abs/1508.01991

[16] Charles R. Jennings. 2013. Social and economic characteristics as determinants of residential fire risk in urban neighborhoods: A review of the literature. *Fire Safety Journal* 62 (2013), 13 – 19. https://doi.org/10.1016/j.firesaf.2013.07.002 Special Issue on Spatial Analytical Approaches in Urban Fire Management.

[17] Renhe Jiang, Xuan Song, Zipei Fan, Tianqi Xia, Quanjun Chen, Qi Chen, and Ryosuke Shibasaki. 2018. Deep ROI-Based Modeling for Urban Human Mobility Prediction. *IMWUT '18* 2, 1 (2018), 14:1–14:29. https://doi.org/10.1145/3191746

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

[19] John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. (2001).

[20] Liang Liu, Wu Liu, Yu Zheng, Huadong Ma, and Cheng Zhang. 2018. Third-Eye: A Mobilephone-Enabled Crowdsensing System for Air Quality Monitoring. *IMWUT '18* 2, 1 (2018), 20:1–20:26. https://doi.org/10.1145/3191752

[21] Yan Liu, Bin Guo, Nuo Li, Jing Zhang, Jingmin Chen, Daqing Zhang, Yinxiao Liu, Zhiwen Yu, Sizhe Zhang, and Lina Yao. 2019. DeepStore: An Interaction-Aware Wide&Deep Model for Store Site Recommendation With Attentional Spatial Embeddings. *IEEE Internet of Things Journal* 6, 4 (2019), 7319–7333. https://doi.org/10.1109/JIOT.2019.2916143

[22] Michael Madaio, Shang-Tse Chen, Oliver L. Haimson, Wenwen Zhang, Xiang Cheng, Matthew Hinds-Aldrich, Duen Horng Chau, and Bistra Dilkina. 2016. Firebird: Predicting Fire Risk and Prioritizing Fire Inspections in Atlanta. In *KDD '16*. ACM, New York, NY, USA, 185–194. https://doi.org/10.1145/2939672.2939682

[23] Michael Madaio, Oliver L Haimson, Wenwen Zhang, Xiang Cheng, Matthew Hinds-Aldrich, Bistra Dilkina, and Duen Horng Polo Chau. 2015. Identifying and prioritizing fire inspections: a case study of predicting fire risk in Atlanta. *Bloomberg Data for Good Exchange, New York, NY, USA* (2015).

[24] Yi Ouyang, Bin Guo, Tong Guo, Longbing Cao, and Zhiwen Yu. 2018. Modeling and Forecasting the Popularity Evolution of Mobile Apps: A Multivariate Hawkes Process Approach. *IMWUT* 2, 4 (2018), 182:1–182:23. https://doi.org/10.1145/3287060

[25] Dongming Qin, Jian Yu, Guojian Zou, Ruihan Yong, Qin Zhao, and Bo Zhang. 2019. A Novel Combined Prediction Scheme Based on CNN and LSTM for Urban $PM_{2.5}$ Concentration. *IEEE Access* 7 (2019), 20050–20059. https://doi.org/10.1109/ACCESS.2019.2897028

[26] Tao Qin, Tie yan Liu, Xu dong Zhang, De sheng Wang, and Hang Li. 2009. Global Ranking Using Continuous Conditional Random Fields. In *Advances in Neural Information Processing Systems 21*, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou (Eds.). Curran Associates, Inc., 1281–1288. http://papers.nips.cc/paper/3402-global-ranking-using-continuous-conditional-random-fields.pdf

[27] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *UAI '09*. 452–461. http://dl.acm.org/citation.cfm?id=1795114.1795167

[28] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing Personalized Markov Chains for Next-basket Recommendation. In *WWW'10*. 811–820. http://doi.acm.org/10.1145/1772690.1772773

[29] Bhavkaran Singh Walia, Qianyi Hu, Jeffrey Chen, Fangyan Chen, Jessica Lee, Nathan Kuo, Palak Narang, Jason Batts, Geoffrey Arnold, and Michael Madaio. 2018. A Dynamic Pipeline for Spatio-Temporal Fire Risk Prediction. In *KDD '18*. ACM, 764–773.

[30] Xishun Wang, Minjie Zhang, and Fenghui Ren. 2018. Sparse Gaussian Conditional Random Fields on Top of Recurrent Neural Networks. In *AAAI'18*. 4219–4226. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16319

[31] Matt Wytock and J. Zico Kolter. 2013. Sparse Gaussian Conditional Random Fields: Algorithms, Theory, and Application to Energy Forecasting. In *ICML'13*. 1265–1273. http://jmlr.org/proceedings/papers/v28/wytock13.html

[32] Shi Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*. 802–810.

[33] Dan Xu, Wanli Ouyang, Xavier Alameda-Pineda, Elisa Ricci, Xiaogang Wang, and Nicu Sebe. 2018. Learning Deep Structured Multi-Scale Features using Attention-Gated CRFs for Contour Prediction. *CoRR* abs/1801.00524 (2018). arXiv:1801.00524 http://arxiv.org/abs/1801.00524

[34] Fengli Xu, Tong Xia, Hancheng Cao, Yong Li, Funing Sun, and Fanchao Meng. 2018. Detecting Popular Temporal Modes in Population-scale Unlabelled Trajectory Data. *IMWUT'18* 2, 1, Article 46 (March 2018), 25 pages. https://doi.org/10.1145/3191778

[35] Fengli Xu, Pengyu Zhang, and Yong Li. 2016. Context-aware Real-time Population Estimation for Metropolis. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 1064–1075. https://doi.org/10.1145/2971648.2971673

[36] Fei Yi, Zhiwen Yu, Fuzhen Zhuang, and Bin Guo. 2019. Neural Network based Continuous Conditional Random Field for Fine-grained Crime Prediction. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*. 4157–4163. https://doi.org/10.24963/ijcai.2019/577

[37] Fei Yi, Zhiwen Yu, Fuzhen Zhuang, Xiao Zhang, and Hui Xiong. 2018. An Integrated Model for Crime Prediction Using Temporal and Spatial Factors. In *ICDM '18*. 1386–1391. https://doi.org/10.1109/ICDM.2018.00190

[38] Xiuwen Yi, Junbo Zhang, Zhaoyuan Wang, Tianrui Li, and Yu Zheng. 2018. Deep Distributed Fusion Network for Air Quality Prediction. In *KDD'18*. ACM, New York, NY, USA, 965–973. https://doi.org/10.1145/3219819.3219822

[39] Huichu Zhang, Yu Zheng, and Yong Yu. 2018. Detecting Urban Anomalies Using Multiple Spatio-Temporal Data Sources. *IMWUT'18* 2, 1, Article 54 (March 2018), 18 pages. https://doi.org/10.1145/3191786

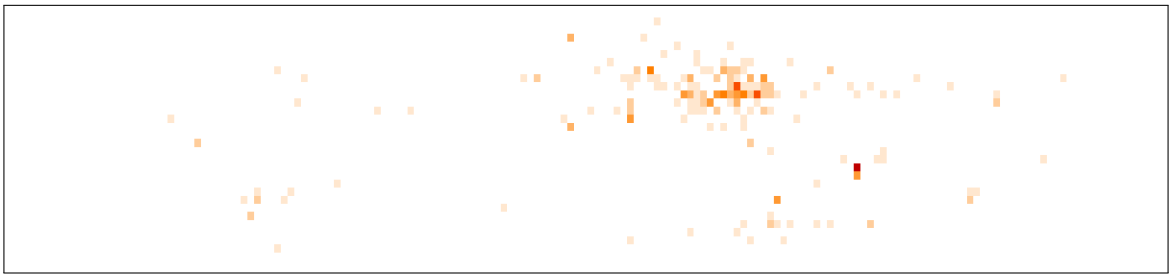
[40] Junbo Zhang, Yu Zheng, Dekang Qi, Ruiyuan Li, Xiuwen Yi, and Tianrui Li. 2018. Predicting citywide crowd flows using deep spatio-temporal residual networks. *Artificial Intelligence* 259 (2018), 147 – 166. https://doi.org/10.1016/j.artint.2018.03.002

[41] Junbo Zhang, Yu Zheng, Junkai Sun, and Dekang Qi. 2019. Flow Prediction in Spatio-Temporal Networks Based on Multitask Deep Learning. *IEEE TKDE* (2019), 1–1. https://doi.org/10.1109/TKDE.2019.2891537

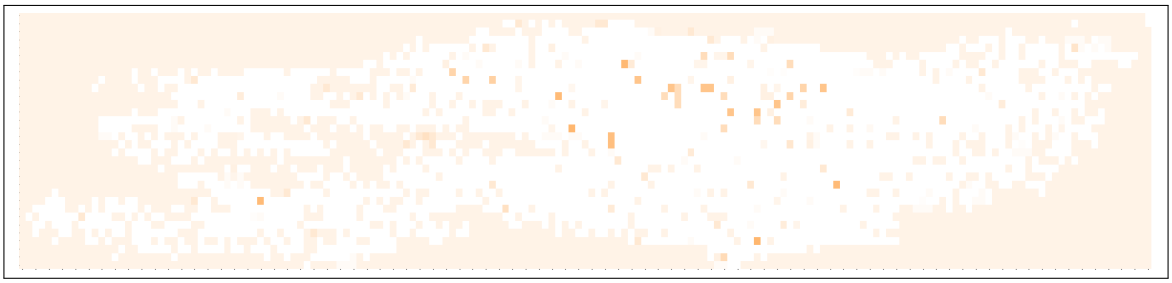
[42] Xiangyu Zhao and Jiliang Tang. 2017. Modeling Temporal-Spatial Correlations for Crime Prediction. In *CIKM'17*. ACM, New York, NY, USA, 497–506. https://doi.org/10.1145/3132847.3133024

[43] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. 2013. U-Air: when urban air quality inference meets big data. In *KDD '13*. 1436–1444. https://doi.org/10.1145/2487575.2488188

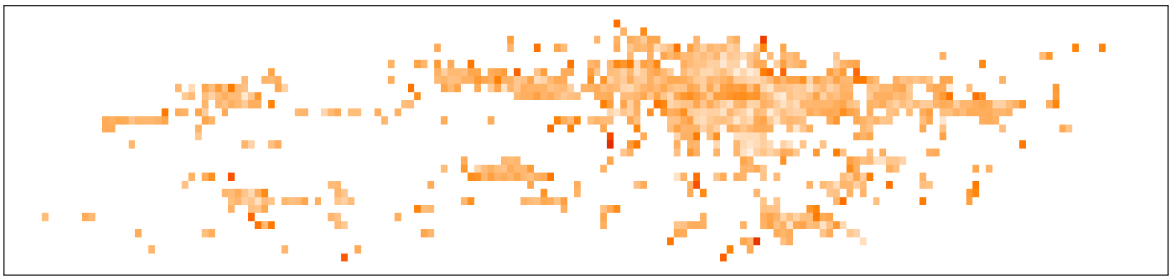
A VISUALIZATION OF RESULTS



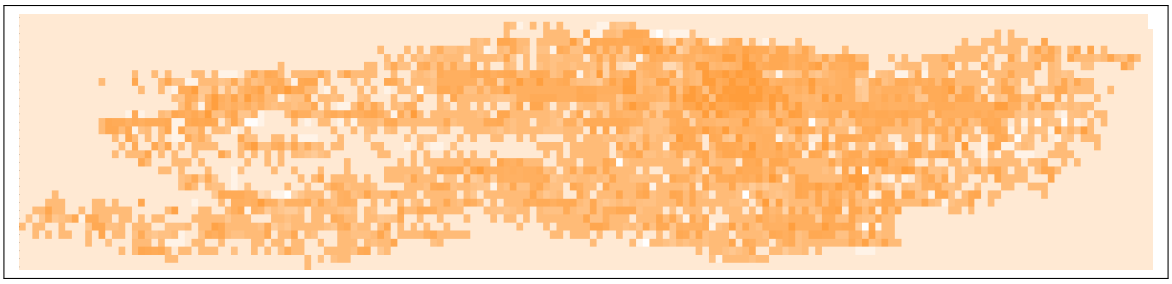
(a) Groundtruth in Jan. 2018



(b) LASSO

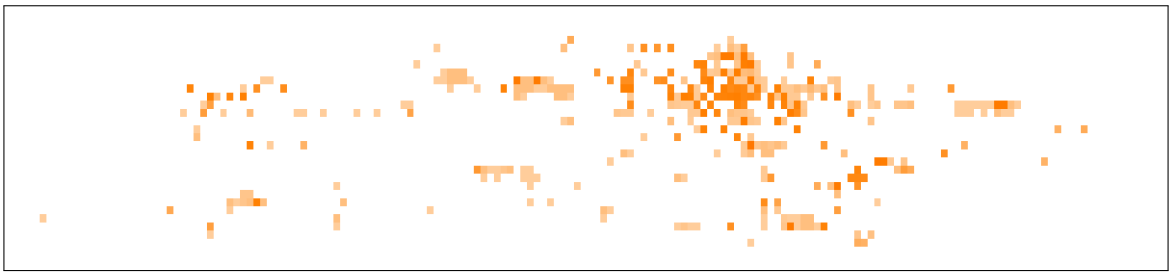


(c) CRF

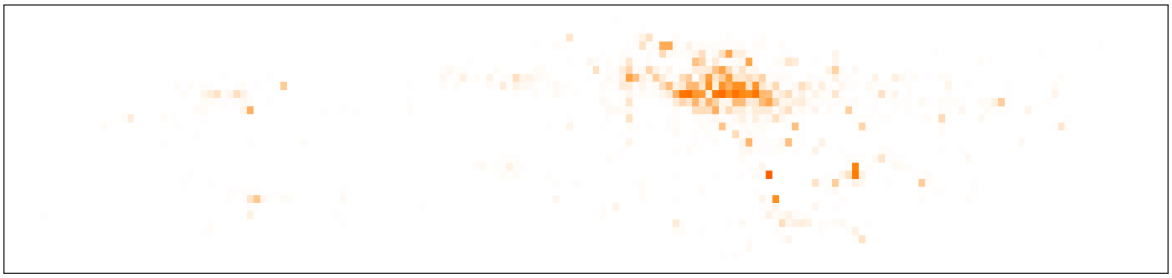


(d) SVM

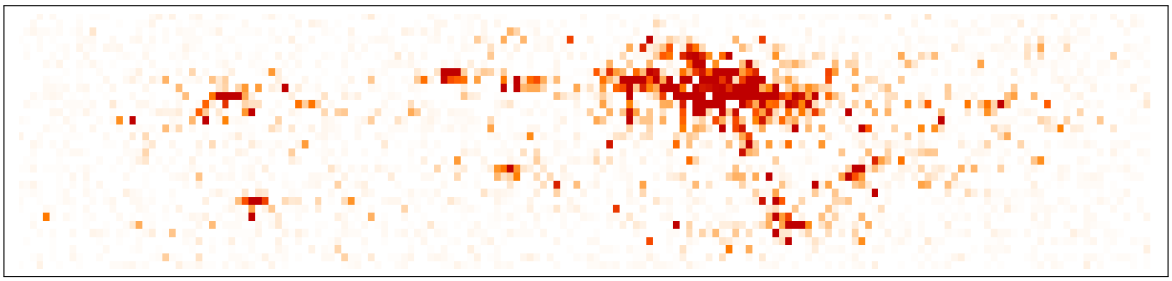
Fig. 17. Visualization of forecasting results



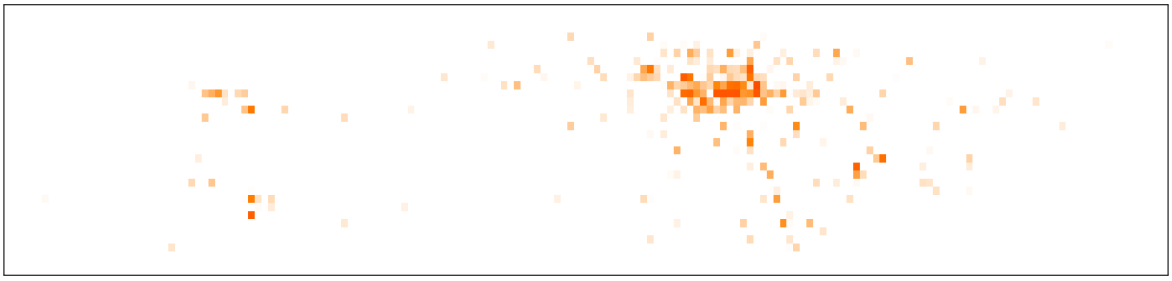
(e) LR



(f) GRU-LR



(g) DeepST



(h) NeuroFire

Fig. 17. Visualization of forecasting results