A Deep Spatio-temporal Residual Network Model for Commercial Activeness Prediction

Ping Liang, DongjinYu*, Xiaoxiao Sun
School of Computer Science and Technology
Hangzhou Dianzi University
Hangzhou, China
liangpingprivate@gmail.com, yudj@hdu.edu.cn, sunxiaoxiao@hdu.edu.cn

***Abstract*—Activeness of regional stores represents the evolvement of their corresponding commercial districts, whose prediction helps practitioners grasp the trend of commercial development and provides support for urban layout. Traditional prediction methods, however, mostly rely on time series analysis and cannot model the complex nonlinear space-time relationship of commercial development as a geographic phenomenon. Inspired by the outstanding performance of deep learning in the field of image and video processing, this paper proposes a deep spatio-temporal residual network model (DSTRN) for commercial activeness prediction using online reviews and check-in records of stores. Specifically, our model includes a spatial dimension that employs local CNN to capture the spatial relation of surrounding commercial districts, and a temporal dimension that applies 3D convolutions and LSTM to deal with the temporal characteristics of commercial activeness. Meanwhile, a residual network is introduced to eliminate gradient vanishing and exploding caused by depth increasing of neural networks. In particular, the recent variations (e.g., sequential changes) and periodic variations (e.g., seasonal changes or holiday effects) of commercial development are both taken into consideration in the model for better prediction. Experiments on public Yelp datasets of Toronto from 2013 to 2018 demonstrate that DSTRN vastly outperforms other approaches and reduces the mean square error by 51.2%, 57.5% and 8.5% compared to Historical average (HA), Autoregressive integrated moving average (ARIMA) and XGBoost, respectively.**

Keywords-Commercial activeness prediction, Commercial district, Spatio-temporal analysis, Deep learning, Yelp

I. INTRODUCTION

Urban commercial districts are key functional areas in cities with high density of shopping malls, restaurants, entertainment facilities and other commercial entities [1]. Activeness of these entities in turn represents the evolvement of their corresponding commercial districts, whose prediction could help practitioners grasp the trend of business development to achieve commercial success and help the government to make accurate urban layout policies.

In recent years, the concepts of big data and smart city have been put forward, which introduce a new idea for predicting commercial activeness based on the massive amount of data collected from different aspects of cities. On the other hand, with the rise of Yelp and other online review apps, a great number of reviews and check-in records of various commercial entities given by visitors and consumers become available. Since business performance, such as revenues and profits, of individual entities are commercial secrets with only limited access, how to employ these publicly available online data to predict the variations of commercial activeness is a topic with practical significance [2]. Traditionally, time series prediction methods such as autoregressive model (ARIMA) are widely applied in commercial prediction field [3,4], which however usually fail to capture the complex non-linear spatio-temporal variations of the commercial activeness as a geographic phenomenon.

To address the above problems, we propose a novel model to predict commercial activeness as a spatio-temporal problem. Our approach is motivated by the outstanding performance of deep learning techniques on handling non-linear relations, more specifically on the prediction of spatio-temporal scenarios such as air pollution prediction [5] and taxi demand scheduling [6]. According to the first law of geography "Near things are more related than distant things", commercial entities in the same district affect the commercial activeness of each other. In order to capture these spatial relations, we apply local Convolutional Neural Networks (local CNN) on commercial activeness prediction, which ensure that close spatial relations are captured and remote spatial relations are excluded. During analysis of the Yelp datasets, we notice that the review and check-in data show periodic changes in some commercial entities. For example, the number of reviews and check-ins of ice cream stores becomes significantly larger in summer than in winter; more customers visit shopping malls in holidays than in workdays. In order to reflect this reality, we adopt the 3D convolutions to extract the recent variations and periodic rules, respectively, and employ Long Short-Term Memory (LSTM) model to synthesize them as the temporal characteristics for commercial activeness prediction.

In summary, the main contributions of this paper are summarized as follows:

- We propose a Deep Spatio-Temporal Residual Network model (DSTRN) to predict the activeness of commercial districts based on online reviews and check-in records of commercial entities.

DOI reference number: 10.18293/SEKE2020-076

*Corresponding author

- We design a spatial dimension in DSTRN that employs local CNN to capture the spatial relations of surrounding commercial districts, and a temporal dimension that applies 3D convolutions and LSTM to deal with the temporal characteristics of commercial activeness (including periodic and recent variations).
- Experiments on public Yelp datasets demonstrate that DSTRN vastly outperforms HA, ARIMA and XGBoost by 51.2%, 57.5% and 8.5% in terms of the mean square error, respectively.

The rest of the paper is structured as follows. After discussing related work in Section 2, we introduce some basic definitions and raise our problem in Section 3. Section 4 presents our prediction model in detail. Experiment results are given in Section 5. Finally, Section 6 concludes the paper and introduces the future work.

II. RELATED WORKS

Space and time are two fundamental dimensions related to all geographic researches. For a long time, spatio-temporal analysis and modeling of geographic parameters have been the main focus of GIScience, such as urban changes [7], land resources utilization [8], and environmental issues [9]. Commercial activeness, which changes with the aggregation and evolution of commercial districts, is also a geographic issue affected by complex space and time factors. Traditional studies on this issue are generally done through investigating local conditions, which is unsustainable and relies too much on field survey [10]. Recent studies attempt to explore the utilities of big data, such as social and review data generated from mobile apps (e.g., Twitter, Yelp, etc.), in commercial activeness prediction. For example, Yang et al. [11] applied a clustering algorithm to aggregate commercial districts based on multiple online data and used a linear model to predict commercial activeness. Hu et al. [2] presented a raster transformation model of check-in data from Internet to analyze commercial district. Wang et al. [12] proposed an approach to the business failure prediction with mobile location-based check-ins. These studies, however, did not recognize the variations of commercial activeness as a non-linear spatio-temporal issue, which led to relatively low accuracy on the prediction results.

Deep learning, which has been vastly applied in the field of image and video processing, is found to be able to handle complex non-linear relations and further to complete spatio-temporal prediction. He et al. [13] put forward a multi-view ensemble neural network to predict commercial hotness, and in some sub-neural networks they introduced CNN. Zhang et al. [14] presented a deep learning model with CNN to predict urban congestion. In these studies, however, the entire research area was input into CNN as one image, which failed to capture the local relations among the surrounding areas and falsely included irrelevant relations of remote entities. In addition, Ji et al. [15] demonstrated that 3D convolution can perceive not only spatial features, but also temporal features compared with 2D convolution in the field of video. Recently, LSTM has been applied to solve spatio-temporal issue due to its outstanding ability in capturing temporal relations. Chen et al. [16] applied LSTM to forecast urban housing price and Kong et al. [17] utilized LSTM to forecast urban power load. Since LSTM cannot reflect the spatial relations, researchers thought that combining CNN and LSTM can capture both spatial and temporal characteristics. For example, Huang et al. [5] applied the CNN-LSTM model to predict air particulate matter (PM2.5). However, due to the complex model architecture, the depths of deep learning network increased sharply, which led to gradient vanishing and exploding and reduced the effectiveness of capturing spatio-temporal relations [18]. On the other hand, LSTM alone cannot capture both the periodic temporal relations (e.g., seasonal changes or holiday effects) in spatio-temporal modeling, which is of vital importance in long-term prediction [14].

In conclusion, commercial activeness prediction is a complex and non-linear geographic issue, which contains both spatial and temporal variations. But until now, these two characteristics have not been adequately extracted at the same time. Furthermore, the temporal features of commercial activeness should be divided into periodic and recent dimensions. Unfortunately, existing methods, even those based on deep learning, however, cannot capture these two dimensions simultaneously and effectively. Therefore, in this paper, we are dedicated to proposing a deep learning model based on local CNN, 3D convolution, LSTM and residual network for commercial activeness prediction.

III. PRELIMINARIES

In this section, we presents the preliminaries which helps understand the model we presented.

A. *Definitions*

Definition 1 (Commercial District). The aggregation of commercial entities in space forms a *Commercial District*, which has a diffusion and radiation effect on its surrounding space.

Definition 2 (Commercial Activeness). The total number of reviews and check-ins from one commercial entity is defined as its *Commercial Activeness*. Similarly, the *Commercial Activeness* of one commercial district is the sum of reviews and check-ins of all commercial entities in this commercial district, defined as follows:

$$y = Sum_{\#Review} + Sum_{\#Checkin} \tag{1}$$

B. *Problem Definition*

Commercial Activeness Prediction: Given the commercial activeness of all grids in the commercial district before a given time t (including t), the problem is to predict the commercial activeness of any grid (i, j) at the time of $t + 1$, defined as:

$$y_{t+1}^{ij} = \mathcal{F}\left(y_{t_{start}}^{ij}, \dots, y_t^{ij}\right) \tag{2}$$

where i and j represent a grid in the commercial district on 2D coordinate, t_{start} represents the time when the earliest data are used as input and $\mathcal{F}$ is the prediction model.

C. *3D convolution*

3D convolution is proved to be effective to capture spatio-temporal features in the field of video convolution [15]. When 2D convolution is applied on video identification, several contiguous frames in the video are treated as multiple channels in the image, which reduces the tensor dimension and makes the network less sensitive to temporal features. However, 2D convolution applies a 2D filter, which only captures spatial features. 3D convolution, by contrast, employs a 3D filter and the increased dimension helps capture the temporal information from contiguous time spans.

D. *Residual Network*

Since the ability of only one convolution layer to capture information is limited, we try to capture depth-dimension information by increasing the numbers of convolution layers. However, training deeper neural networks usually leads to gradient vanishing and exploding. To overcome this problem, we introduce a residual neural network into our model. Residual neural network contains several residual units, each one of which includes two convolution layers and one batch normalization layer. One residual unit is defined as:

$$X^{l+1} = \mathcal{F}_{res}\,(X^{l}) + X^{l} \tag{3}$$

where X^{l} and X^{l+1} are the input and output of the l^{th} residual unit, and $\mathcal{F}_{res}$ is a residual function.

IV. PREDICTION OF COMMERCIAL ACTIVENESS

In this section, we introduce the details of DSTRN, i.e., our prediction model $\mathcal{F}$. As is shown in Fig.1, the framework is mainly composed of two parts, which are used to capture spatial correlation and temporal correlation, respectively. As for temporal correlation, both periodic variations and recent variations are considered.

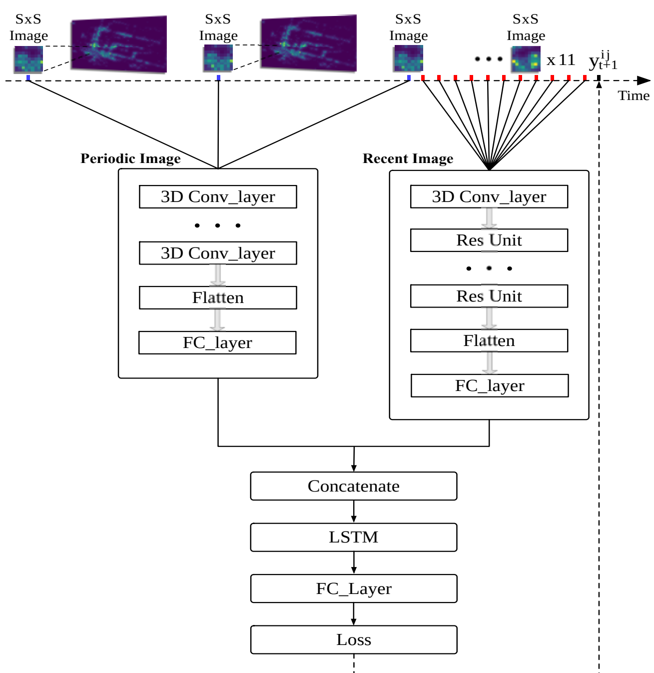


Figure 1. The framework of deep spatio-temporal residual network.

A. *Preprocessing*

Commercial entities are spatially discrete, which are not conducive to the capture of their spatial correlations. To eliminate this obstacle, we rasterize the commercial district into a series of grids and sum the counts of reviews and check-ins about the business entities located inside a certain grid to indicate the commercial activeness of the corresponding grid as Fig.2 shows. We set the time span as one month and obtain a time sequence expressed as $\Gamma = \{T_0, T_1, T_2, T_3, \dots, T_t\}$. For each time slot T_t, we rasterize the commercial activeness and obtain a set of grid activeness represented as $\xi^{t} = \{P^{t}_{00}, P^{t}_{01}, P^{t}_{02}, \dots, P^{t}_{ij}\}$.

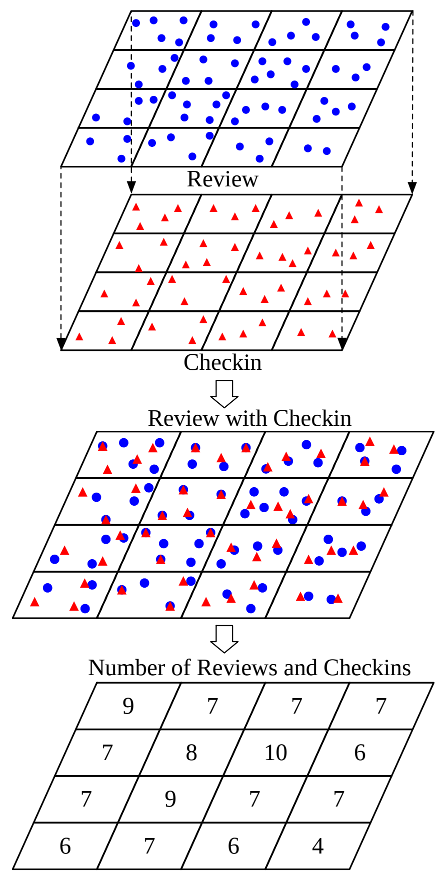


Figure 2. Superposition of reviews and check-ins to indicate commercial activeness.

B. *Spatial Information Extraction*

As is mentioned above, in previous spatio-temporal studies, the whole research area (i.e., the commercial district) is input into CNN as image with each grid having its commercial activeness as its value. To focus on correlations among the surrounding areas and eliminate irrelevant relations of remote areas, we introduce local CNN to extract spatial information of commercial activeness. The following steps demonstrate the process of image segmentation as is shown in Fig.3:

1) For each time slot T_t, we place the commercial district in the first quadrant of a 2D coordinate system.

2) We extract grid P_{ij} and its surrounding grids as an S ∗ S image starting from the origin of coordinates, where S is an odd number to maintain P_{ij} in the center of image.

3) We then move P_{ij} one grid to the right and repeat the segmentation operation until the right edge of the last segmented image reaches the right border of the research area.

4) Afterwards, we move P_{ij} one grid up each time until the upper edge of the segmented image reaches the upper border of the research area.

So far, we obtain an image tensor set expressed as $Y_t^{ij} \in R^{S*S}$, in which every image has commercial activeness as pixel value of each grid. The local CNN takes Y_t^{ij} as an input image to convolutional layers, which is defined as follows:

$$Y_t^{ij,k+1} = f\left(Y_t^{ij,k} * W_t^k + b_t^k\right) \tag{4}$$

where ∗ denotes the operation of CNN and f is an activation function. $Y_t^{ij,k}$ is the input of the k^{th} CNN layer. W_t^k and b_t^k are learnable parameter sets. Since the task is to predict the commercial activeness of the central grid within an S ∗ S image, our model does not involve any subsampling and pooling operations

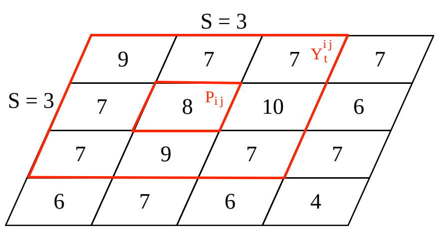


Figure 3. Image segmentation for local CNN (when S is 3).

C. Temporal Correlation Extraction

The temporal feature extraction of commercial activeness applies a 3D convolution, in which the image tensor Y_t^{ij} in each time span is regarded as a video frame. 3D convolution is in fact an increased-dimension 2D convolution. So local CNN is applied in each 2D dimension of 3D convolution to capture spatial information. Considering the difference between periodic and recent characteristics of commercial activeness, a periodic neural network is designed together with a recent neural networks as Fig.1 shows.

When designing the periodic neural network, several 3D convolutions are applied to extract periodic spatio-temporal information. For each 3D convolution, the activation function is represented as:

$$\mathcal{Y}_t^{ij,k+1} = f\left(\mathcal{Y}_t^{ij,k} * W_t^k + b_t^k\right) \tag{5}$$

where ∗ denotes the operation of 3D convolution, $\mathcal{Y}_t^{ij,k}$ is the input of the k^{th} 3D convolution layer, and W_t^k and b_t^k are learnable parameter sets. After several 3D convolution layers, we flatten the output tensor $\mathcal{Y}_t^{ij} \in R^{S*S}$ to a feature vector $v_t^{ij} \in R_t^{S^2}$ for grid (i,j) at time slot t. Finally, a fully connected layer is applied to reduce the length of spatio-temporal feature vector v_t^{ij} , which is defined as:

$$\zeta_t^{ij} = f\left(W_t^{FC^{3d}} v_t^{ij} + b_t^{FC^{3d}}\right) \tag{6}$$

where f is an activation function, and $W_t^{FC^{3d}}$ and $b_t^{FC^{3d}}$ are learnable parameters sets. So far, we get a periodic spatio-temporal information vector $\zeta_{t\_periodic}^{ij} \in R^l$, where l means length of the vector. Thus, the periodic data in p periods generate p $\zeta_{t\_periodic}^{ij}$, which is then integrated into one tensor $\eta_{t\_periodic}^{ij} \in R^{l*p}$.

When designing the recent neural network, we assume that recent correlation is more relevant to the prediction result and hence we input more recent data to our model. Based on this idea, we increase the depth of recent correlation extraction neural networks, i.e., increasing the number of 3D convolution layers. Since increasing layers leads to gradients vanishing, we introduce residual network to extract recent information. As mentioned above, in each residual unit, two 3D convolution layers and one batch normalization layer are employed, together with several residual units, as Fig.1 shows. After a flatten layer and a fully connected layer, we obtain a recent spatio-temporal information tensor $\eta_{t\_recent}^{ij} \in R^{l*q}$, in which q represents the number of recent time slots.

In order to automatically assign different weights to recent correlation and periodic correlation, LSTM network is applied in our model. The concatenation of the two tensors $\eta_{t\_periodic}^{ij}$ and $\eta_{t\_recent}^{ij}$ is as follows:

$$\theta_t^{ij} = \eta_{t-period}^{ij} \oplus \eta_{t-recent}^{ij} \tag{7}$$

Here, the tensor θ_t^{ij} contains $p+q$ periods of spatio-temporal information. θ_t^{ij} is then fed into LSTM and an outputting tensor is obtained as δ_t^{ij}.

D. Prediction

For the final prediction, our goal is to obtain all values of grid activeness for time slot $t+1$. Since tensor δ_t^{ij} already contains spatial and temporal correlations, a fully connected layer is applied to calculate the final prediction value $\hat{y}_{t+1}^{ij}$ as follows:

$$\hat{y}_{t+1}^{ij} = f\left(W_t^{FC} \delta_t^{ij} + b_t^{FC}\right) \tag{8}$$

where W_t^{FC} and b_t^{FC} are learnable parameters, and f is the activation function. The final result is normalized to [0,1] and denormalized back to real commercial activeness values. According to the order that local CNN segments images, the predicted values are calculated back to obtain the predicted grid values of grid activeness.

E. Loss Function

Our model is trained through minimizing the value of loss function iteratively, which is defined as the average absolute error (MAE) between the real commercial activeness value and the predicted one as Eq.(9) shows:

$$L_\theta = \frac{1}{m}\sum_{i=1}^{m}\left|y_t^{ij} - \hat{y}_t^{ij}\right| \tag{9}$$

where θ is all learnable parameters in DSTRN, and m is the number of samples.

V. EXPERIMENTS

A. Experimental Settings

1) Datasets: We selected the open dataset[1] from Yelp for the experiments, which includes more than 668,000 reviews and more than 160,000 check-in records of stores in various cities from October 2004 to November 2018. We chose Toronto as the target city to evaluate the performance of the proposed method DSTRN.

After investigation, we selected the data from January 2013 to December 2017 as the training dataset, and the data from 2018 as the testing dataset, and adopted one month as the time span. When testing the prediction results, we used the same month in the previous 3 years and the previous 11 months to predict the commercial activeness in the next time slot. In our experiment, the activeness values smaller than 10 were filtered, which is a common practice in practical applications [6].

2) Parameters Settings: We implemented DSTRN[2] by Keras, which is a fast experimentation neural networks API running on top of TensorFlow. The experiments ran on a cluster with four NVIDIA 1080Ti GPUs. During experiments, we applied min-max normalization on training datasets to normalize input values to [0,1]. After the DSTRN prediction, we reversed the min-max normalization to recover commercial activeness values.

In our experiment, the size of surrounding grids was set to 7 $*$ 7, which corresponds to a 7km$*$7km actual rectangle area. For the periodic part of DSTRN, we used 3 3D convolution layers with 32 filters, of which the size is 3 $*$ 3 $*$ 3. For the recent part of DSTRN, we set the number of residual units to 12, and 32 filters with size of 3 $*$ 3 $*$ 3 were applied in the 3D convolution layers of each residual unit. In all convolution layers, we adopted ReLU as the activation functions. For LSTM, we set the dimensions of hidden state vector as 512. In the final fully connection layer, we adopted Sigmoid as activation function. As for other parameters, the batch size was set to 64, and the learning rate was 0.001. Finally, we applied Adam as optimizer.

3) Evaluation Metrics: We use mean square error (MSE) and mean percentage error (MAPE) as accuracy indicators of the model, which are defined as follows:

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(y_t^{ij} - \hat{y}_t^{ij}\right)^2 \tag{10}$$

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{y_t^{ij} - \hat{y}_t^{ij}}{y_t^{ij}}\right| \times 100 \tag{11}$$

where y_t^{ij} is the observed activeness value, $\hat{y}_t^{ij}$ is the predicted value for grid (i, j) at time slot t and m is the number of samples.

4) Comparison Methods: We compared our model with three other methods, i.e., Historical average (HA), Autoregressive integrated moving average (ARIMA), and XGBoost.

- **Historical average (HA):** A traditional prediction method, in which the predicted value is the average value of previous historical commercial activeness values.
- **Autoregressive integrated moving average (ARIMA):** A traditional time series analysis method, which does not consider the variations of other relevant random variables.
- **XGBoost:** A gradient boosting tree model, which has gained widely popularity recently after many winning teams of competitions used it.

B. Experiment Results

In this section, we demonstrate the experimental results of DSTRN compared with other methods in the first experiment and we further try to adjust parameters in our model to see their influence on the model performance (i.e., size of surrounding grids, the numbers of recent months and the numbers of residual units) in the second experiment.

1) Model Performance: Figure 4 shows that DSTRN achieves the lowest MSE (i.e., 26.25) and the lowest MAPE (i.e., 16.20) among all methods. Compared to the second-best model (i.e., XGBoost), DSTRN improves 8.5% in MSE and 26.9% in MAPE, respectively. Compared to traditional methods such as HA and ARIMA, machine learning and deep learning models obviously have better performances in terms of MSE and MAPE.

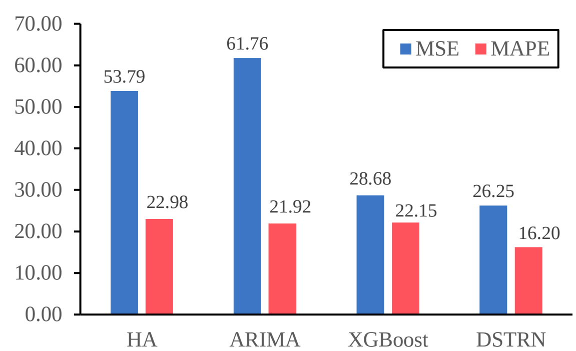


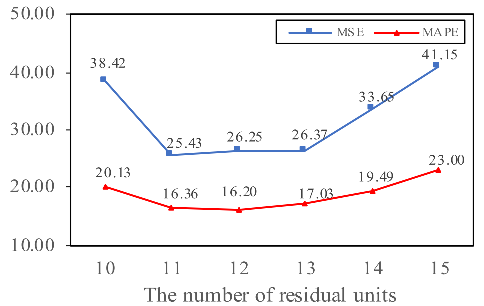
Figure 4. Comparision of different methods.

2) Parameter Influence: In the second experiment, we explore how the number of residual units influences the accuracy of our model as shown in Fig.5a. When the number of residual units ranges between 11 and 13, MSE and MAPE are relatively stable. Otherwise, MSE rises significantly due to the lack of deep mining of commercial activeness. However, when the number of residual units increases to a certain amount, the gradients vanishes gradually. Since MAPE better reflects the overall prediction accuracy of the model, we finally apply 12 residual units in our model.

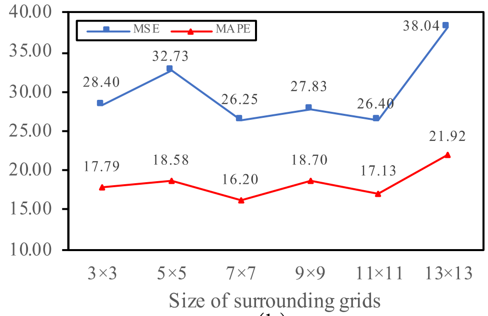
In addition, because the size of surrounding grids determines the input image size of local CNN, we also tried to figure out which size is best for the model. As shown in Fig.5b, when the size is selected as 7 $*$ 7, MSE and MAPE both obtain their optimal values. When we increase the size to 13 $*$ 13, MSE and MAPE rise significantly. The reason is that several unrelated locations are included in CNN and hence reduces the precision

[1]https://yelp.com/dataset
[2]http://dbsi.hdu.edu.cn/CommercialActiveness/

of the model. MSE and MAPE also increase slightly as the size decreases to 5 ∗ 5 for not considering enough surrounding relations.



(a)



(b)

Figure 5. Comparison of DSTRN with different parameters

VI. CONCLUSION

In this paper, we propose a model called DSTRN which employs online reviews and check-in records to predict commercial activeness by month. Specifically, our model includes a spatial dimension that employs local CNN to capture the spatial relation of surrounding commercial districts and a temporal dimension that applies 3D convolutions to deal with the temporal characteristics of commercial activeness. Considering the obvious periodic and recent patterns of commercial activeness, DSTRN handles periodic and recent temporal dimensions simultaneously and applies LSTM to combine the both. In addition, to avoid the gradients vanishing and exploding caused by increasing of convolution layers, the residual network is applied in DSTRN. Experiments on public Yelp datasets of Toronto from 2013 to 2018 demonstrate that DSTRN vastly outperforms other methods.

For future optimization, we plan to add more related information about commercial activeness, such as the layout of subways, as the input of the prediction other than just reviews and check-in records. Besides, the semantics of reviews should also be considered in the future to improve the performance of the model.

ACKNOWLEDGMENTS

This work was partially supported by National Natural Science Foundation of China (No. 61472112, No. 61702144), Key Science and Technology Project of Zhejiang Province of China (No. 2017C01010), and Natural Science Foundation of Zhejiang Province of China (No. LQ20F020017). The authors would also like to thank anonymous reviewers who made valuable suggestions to improve the quality of the paper.

REFERENCES

[1] Wang, Fang, Li, Yan, and Gao, Xiaolu. "A SP survey-based method for evaluating environmental performance of urban commercial districts: A case study in Beijing." Habitat International 53 (2016): 284-291.

[2] Hu, Qingwu,Wang, Ming, and Li, Qingquan. "Urban hotspot and commercial area exploration with check-in data." Acta Geodaetica et Cartographica Sinica 43.3 (2014): 314-321.

[3] Nath, Bhola, D. S. Dhakre, and Debasis Bhattacharya. "Forecasting wheat production in India: an ARIMA modelling approach." Journal of Pharmacognosy and Phytochemistry 8.1 (2019): 2158-2165.

[4] Zhu, Bangzhu, and Julien Chevallier. "Carbon price forecasting with a hybrid arima and least squares support vector machines methodology." Pricing and Forecasting Carbon Markets. Springer, Cham, 2017. 87-107.

[5] Huang, Chiou-Jye, and Ping-Huan Kuo. "A deep cnn-lstm model for particulate matter (PM2. 5) forecasting in smart cities." Sensors 18.7 (2018): 2220.

[6] Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Ye, J., Chuxing, D., and Li, Z. "Deep multi-view spatial-temporal network for taxi demand prediction." Thirty-Second AAAI Conference on Artificial Intelligence. 2018.

[7] Clarke, Keith C., and Leonard J. Gaydos. "Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore." International journal of geographical information science 12.7 (1998): 699-714.

[8] Wrenn, Douglas H., and Abdoul G. Sam. "Geographically and temporally weighted likelihood regression: Exploring the spatiotemporal determinants of land use change." Regional Science and Urban Economics 44 (2014): 60-74.

[9] Liu, Y., Zheng, Y., Liang, Y., Liu, S., Rosenblum, D.S.: Urban water quality prediction based on multi-task multi-view learning. Proceedings of the 25th International Joint Conference on Artificial Intelligence. pp. 2576-2582 (2016).

[10] Mandhachitara, Rujirutana, and Randall Shannon. "The Formation and Sustainability of same Product Retail Store Clusters in A Modern Mega City." Tijdschrift voor economische en sociale geografie 107.5 (2016): 567-581.

[11] Yang, S., Wang, M., Wang, W., Sun, Y., Gao, J., Zhang, W., and Zhang, J. "Predicting commercial activeness over urban big data." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1.3 (2017): 1-20.

[12] Wang, lei, Gopal, Ram, Shankar, Ramesh, and Pancras Joseph. "On the brink: Predicting business failure with mobile location-based checkins." Decision Support Systems 76 (2015): 3-13.

[13] He, Zhiyuan, and Su Yang. "Multi-view Commercial Hotness Prediction Using Context-aware Neural Network Ensemble." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2.4 (2018): 1-19.

[14] Zhang, Junbo, Yu Zheng, and Dekang Qi. "Deep spatio-temporal residual networks for citywide crowd flows prediction." Thirty-First AAAI Conference on Artificial Intelligence. 2017.

[15] Ji, Shuiwang, Xu, Wei, Yang, Ming, and Yu, Kai. "3D convolutional neural networks for human action recognition." IEEE transactions on pattern analysis and machine intelligence 35.1 (2012): 221-231.

[16] Chen, Xiaochen, Wei, Lai, and Xu, Jiaxin. "House Price Prediction Using LSTM." arXiv preprint arXiv:1709.08432 (2017).

[17] Kong, Weicong, Dong, Zhaoyang, Jia, Youwei, Hill, D., Yan, Xu and Zhang, Yuan. "Short-term residential load forecasting based on LSTM recurrent neural network." IEEE Transactions on Smart Grid 10.1 (2017): 841-851.

[18] He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, and Sun, Jian. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.