Forecasting Taxi Demands with Fully Convolutional Networks and Temporal Guided Embedding

Doyup Lee[1,2], **Suehun Jung**[2], **Yeongjae Cheon**[2], **Dongil Kim**[3], **Seungil You**[3]
[1]Department of Creative IT Engineering, Pohang University of Science and Technology
[2]Kakao Brain
[3]Kakao Mobility
zzehqlzz@postech.ac.kr, {noah.nobel, yeongjae.cheon}@kakaobrain.com
{bob033, sean.you}@kakaomobility.com

Abstract

Learning complex spatiotemporal patterns is a key to predict future taxi demand volumes. We propose temporal guided networks (TGNet), which is an efficient model architecture with fully convolutional networks and temporal guided embedding, to capture spatiotemporal patterns. Existing approaches use complex architectures, historical demands (day/week/month ago) to capture the recurring patterns, and external data sources such as meteorological, traffic flow, or texture data. However, TGNet only uses fully convolutional networks and temporal guided embedding without those external data sources. In this study, only pick-up and drop-off volumes of NYC-taxi dataset are used to utilize the full potential of the hidden patterns in the historical data points. We show that TGNet provides notable performance gains on a real-world benchmark, NYC-taxi dataset, over previous state-of-the-art models. Finally we explain how to extend our architecture to incorporate external data sources.

1 Introduction and Related Works

On-demand ride hailing platforms, such as Uber, Didi, Lyft, and Kakao T, fulfill more than millions ride hailing demands per day and become essential part of the urban life style. Short-term demand forecasting is essential in those platforms since dynamic pricing relies on real-time demand forecast and dispatch system can relocate drivers to high demand areas. For example, by adjusting price in a high demand region, the potential short term supply and demand imbalances can be controlled.

To predict future demand volumes in each region, a predictive model must learn complex spatiotemporal correlations, because a demand volume of each region is highly correlated with a previous demand volume (temporal pattern) and demand volumes of connected region are also highly correlated (spatial pattern) [4, 9]. To learn a temporal pattern, a statistical approach, such as auto-regressive integrated moving average (ARIMA), can be used in each region, but this approach may fail to capture the spatial correlations. Vector auto-regressive (VAR) model may be used to learn spatial patterns, but this model is known to have limitation on the high dimensional data [1].

To overcome the aforementioned challenges and learn spatiotemporal patterns, recent studies use deep neural networks to predict future demand volumes and show notable performance gains over the traditional approaches. To learn a spatial patterns, [12, 11] formed a grid over the region of interests, and aggregated the data of interest, *e.g.*, crowd flows, as pixel values. After this transformation, we can think a demand pattern as an image, and convolutional networks can be employed to model local spatial correlations.

Workshop on Modeling and Decision-Making in the Spatiotemporal Domain, 32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

DMVST-Net [9] modeled spatial, temporal, and semantic contexts by combining local CNN, LSTM, and graph convolutional networks (GCN) all together. STDN [8], which consists of local CNN, LSTM, and the temporal attention, predicted future volumes of taxi demand and used external data sources such as traffic flow and meteorological data. STDN also showed state-of-the-art performances on NYC-taxi dataset. In addition, previous approaches [11, 9, 8] used periodic/seasonal histories (days/weeks ago) to capture recurrent patterns in the time series data.

Incorporating external data sources can also improve the forecasting quality. In [12, 11, 9, 8], meteorological data is commonly used. [6] used temporal, spatial, meteorological, event, and their combination data to predict future taxi demands. Traffic in/out flow [8] and textual data from web sites [5] are also used to forecast demands.

In this paper, we propose an efficient model framework (TGNet), fully convolutional networks with temporal guided embedding. Our proposed architecture does not use complex model and does not require external data sources to achieve the state-of-the-art results, in contrast to the previous state-of-the-art models. We show that our new architecture is enough to model complex spatiotemporal correlation within a dataset to achieve the best performance. We believe that our proposed architecture can provide competitive baseline results on many spatiotemporal datasets.

The main contributions in this paper is as follows. First of all, TGNet has a simple architecture to learn complex spatiotemporal patterns in taxi demands unlike previous approaches that combine complex models and external data sources. Secondly, we show notable performance gains using temporal guided embedding, which helps the model extract hidden representations conditioned on temporal context over the traditional approaches. Thirdly, we show competitive performance on NYC-taxi dataset without utilizing external data sources compared to the current state-of-the-art results that uses the external data sources. Lastly, we also show how to extend TGNet to incorporate external data sources.

2 Dataset and Methods

2.1 Dataset and Preprocessing

We uses NYC-Taxi dataset, which contains taxi trip records of NYC in 2015, from 01/01/2015 to 03/01/2015. Taxi pick-up and drop-off volumes in NYC are recorded at every 30 minutes interval. In this study, the records in first 40 days data are trained and the remaining records are tested.

In this paper, we split the area of NYC into $I \times J$ grid map. We define the set of non-overlapped region $L = \{L_{11}, ..., L_{IJ}\}$. Also, the set of non-overlapped time intervals is defined by $\mathcal{I} = \{I_1, ..., I_p\}$, whose interval is 30 minutes. A taxi demand r has time stamp and request location $(r.t, r.l)$. Then, the taxi demands at time interval I_t are defined by $X^{(t)} = \{X_{ij}^{(t)} : L_{ij} \in L\}$, where $X_{ij}^{(t)} = |\{r : r.t \in I_t \wedge r.l \in L_{ij}\}|$. In here, $|\cdot|$ denotes the cardinality of the set. In this study, we convert the data into $\mathbf{X}^{(t:t-T)} \in \mathbb{R}^3$, which is an image with $(T+1)$ channels. Here, the colon, :, means channel wise concatenation, that is $\mathbf{X}^{(t:t-1)} = \left\{X^{(t)}; X^{(t-1)}\right\}$.

2.2 Forecasting Problem Formulation

To forecast $t+1$ future volumes at time t, we use adjacent past volumes of taxi requests until $t-T$ as features. Periodic and seasonal inputs are not used to capture recurrent patterns in taxi demands, but temporal guided embedding replaces them. We introduce temporal guided embedding in the next section. The model $\mathcal{F}$ forecasts $t+1$ future volumes only with $X^{(t:t-T)}$ and $\mathcal{E}^{(t:)}$, where $\mathcal{E}^{(t:)}$ is external data available at t. Then, $\hat{X}^{(t+1)} = \mathcal{F}(\mathbf{X}^{(t:t-T)}, \mathcal{E}^{(t:)})$

3 TGNet: Fully Convolutional Networks with Temporal Guided Embedding

3.1 Baseline Model Architecture

Figure 1(a) shows fully convolutional networks to forecast taxi demands. TGNet has symmetric convolution and deconvolution operations with skip connections. Skip connections prevent high level features of deconvolution from losing precise positional information of each pixel after a stack of convolutional and pooling layers. Average pooling is used to aggregate the feature values of receptive area and each subsample represents the characteristic of the area. Dense block contains three convolutional layers with dense skip connections in [2] to help model learn after pooling. Each subsample of average pooling represents the semantic features of its receptive area and the dense block is expected to learn semantic context between large regions. 1×1 convolutional layers are located at the end of model and operate as position-wise feature extractor and regression. They integrate various feature sources to predict future demand volume of whole target regions. It does not consider spatial correlation anymore, but focus on forecast requests with precise localization.

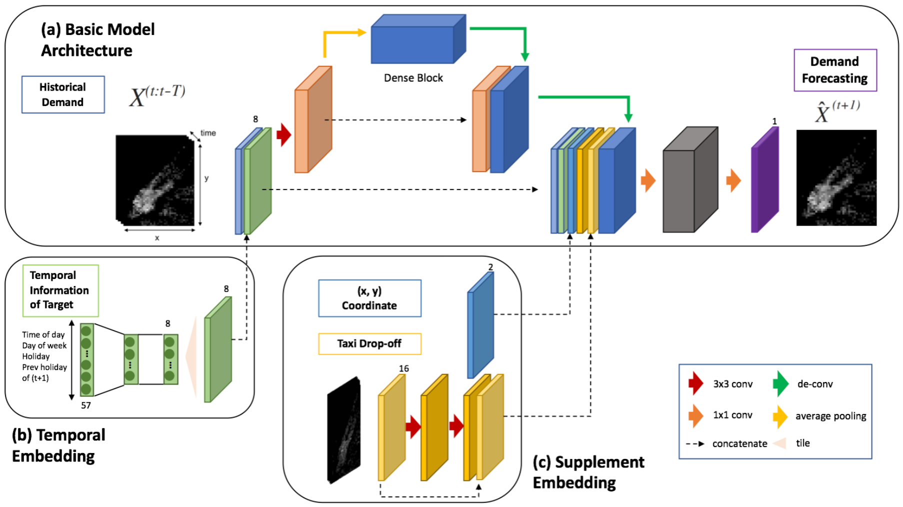

Figure 1: TGNet Model Framework: (a) baseline architecture, (b) temporal guided embedding, and (c) supplement embedding of external data sources

3.2 Temporal Guided Embedding

Traffic data, such as taxi demand, is known to have temporally recurrent patterns, such as periodicity and seasonality. The patterns are different for time-of-day, day-of-week, and holiday or not. To learn these patterns, existing researches use corresponding historical inputs such as demand volumes of a day (week) ago at the same time of day, and show promising results [12, 11, 8].

However, we postulate that this may not be the optimal way to digest such a temporal patterns. For example, when a person is asked to forecast a taxi demand volume, he or she may want to know temporal information, such as time-of-day, day-of-week, and holiday information. If not, it is just pattern matching with input data, without understanding of temporal contexts in taxi demands. This indicates that using temporal information can be much effective to inference the recurring patterns. Thus, we use temporal information as categorical variables: time-of-day (48 dimensional variables for every 30 minutes), day-of-week (7), holiday (1), and previous day of holiday (1).

TGNet learns embedding of temporal information and it is concatenated on demand inputs as temporal guidance. The fully convolutional networks extract hidden representations of complex spatiotemporal patterns in taxi demands, conditioned on temporal embedding. We expect that temporal guided embedding is enough to capture temporal contexts than usage of periodic and seasonal inputs. Note that figuring out the right period and seasonality by looking at partial ACF is often time consuming, which shows the temporal guided embedding's another advantage over the existing approaches.

3.3 Late Fusion with Supplemental Data Sources

In TGNet, features from other data sources can be concatenated to feature maps that are extracted by deconvolutional layers. This type of late fusion is a common approach to combine intra-modality of heterogeneous data sources [10]. In this paper, we encode drop-off information in NYC-taxi dataset and concatenate on position-wise layer. In the future, we expect various types of information can be attached in the same manner.

Table 1: Performances on NYC-Taxi: ConvLSTM [7], DMVST-Net [9], STDN [8], and TGNet.

	ARIMA	ConvLSTM	DMVST -Net	STDN	TGNet: (a)	TGNet: (a)+(b)	TGNet: (a)+(b)+(c)
MAPE	22.21	20.50	17.36	16.30	17.46	**15.65**	**15.45**
RMSE	36.53	28.13	25.71	**24.10**	28.97	25.31	**25.18**
Data sources	-	-	weather	weather, traffic flow	-	-	-
Model details	-	ConvLSTM	CNN, LSTM, GCN	CNN, LSTM, attention	FCN	FCN	FCN
Trainable variables	-	-	-	9,446,274	-	-	1,241,441

4 Experiment Results

4.1 Implementation

The area of NYC is divided into 20×10 grids. Recent historical volumes of taxi demand (4 hours) and drop-off (8 hours) are used. Hyper-parameters of compared models are followed original papers. In TGNet, 32, 128, 128, 128, and 128 filters of 3×3 (de)convolutional layers and two 1x1 convolutional layers with 256 filters are used. Temporal information and drop-off are encoded by 8 and 32 filters, respectively. Adam optimizer [3] with learning rate 0.01 and decay 0.01 are used. Batch size is 128 and the model is trained using early stopping, validated by 20% of training data. After pre-train using L2 loss, the model is optimized by L1 loss function, because it's robust on noises in traffic data [8].

4.2 Numerical Results

In evaluation, we filtered low-value samples less than 11 to obtain a meaningful performance metric [8], since a low demand value is not a concern in a real-world application. Mean absolute percentage error (MAPE) and root of mean squared error (RMSE) are used as performance measures. In table 1, (a), (b), and (c) notate the components in Figure 1. Temporal guided embedding (b) shows notable performance gains of MAPE and RMSE, capturing recurrent patterns in taxi demands without previous periodic and seasonal inputs. TGNet also shows the best MAPE performance and competitive result of RMSE. The results are noteworthy. The state-of-the-art model, STDN, combines local CNN and LSTM with temporal closeness/periodic/seasonal inputs and incorporates external data, such as traffic flow and meteorological data. However, TGNet only contains FCN structure with temporal guided embedding and uses drop-off data. The number of trainable parameters is also 8 times smaller than previous state-of-the-art model, STDN. We conclude that TGNet efficiently captures complex spatiotemporal patterns in taxi demand itself.

5 Conclusion and Discussion

In this paper, we propose temporal guided networks (TGNet), fully convolutional networks with temporal guided embedding to capture spatiotemporal pattern in taxi demands. Especially, temporal guided embedding learns and helps FCN to extract spatiotemporal features conditioned on temporal contexts. Indeed, it provides notable performances gains on forecasting taxi demands on real-world dataset. We also show that TGNet has simple architecture to add various data sources with late fusion.

In the future we will investigate how to encode other context information, including meteorology, articles, and other traffic information. Our study will be extended to analyze atypical events and predict multi-step ahead future, based on TGNet architecture. Especially, uncertainty calibration of both data and model can be important issue of interpretability when abrupt changes or atypical events occur. Finally, we expect that TGNet can be applied to solve other spatiotemporal problems in general.

References

[1] Sumanta Basu. *Modeling and Estimation of High-dimensional Vector Autoregressions*. PhD thesis, The University of Michigan, 2014.

[2] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *CVPR*, volume 1, page 3, 2017.

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

[4] Xiaolei Ma, Zhuang Dai, Zhengbing He, Jihui Ma, Yong Wang, and Yunpeng Wang. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, 17(4):818, 2017.

[5] Filipe Rodrigues, Ioulia Markou, and Francisco C Pereira. Combining time-series and textual data for taxi demand prediction in event areas: a deep learning approach. *Information Fusion*, 2018.

[6] Yongxin Tong, Yuqiang Chen, Zimu Zhou, Lei Chen, Jie Wang, Qiang Yang, Jieping Ye, and Weifeng Lv. The simpler the better: a unified approach to predicting original taxi demands based on large-scale online platforms. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1653–1662. ACM, 2017.

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

[8] Huaxiu Yao, Xianfeng Tang, Hua Wei, Guanjie Zheng, Yanwei Yu, and Zhenhui Li. Modeling spatial-temporal dynamics for traffic prediction. *arXiv preprint arXiv:1803.01254*, 2018.

[9] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye, and Zhenhui Li. Deep multi-view spatial-temporal network for taxi demand prediction. In *AAAI*, pages 2588–2595, 2018.

[10] Amir Zadeh, Minghai Chen, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. Tensor fusion network for multimodal sentiment analysis. *arXiv preprint arXiv:1707.07250*, 2017.

[11] Junbo Zhang, Yu Zheng, and Dekang Qi. Deep spatio-temporal residual networks for citywide crowd flows prediction. In *AAAI*, pages 1655–1661, 2017.

[12] Junbo Zhang, Yu Zheng, Dekang Qi, Ruiyuan Li, and Xiuwen Yi. Dnn-based prediction model for spatio-temporal data. In *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, page 92. ACM, 2016.