

Multiple Areal Feature Aware Transportation Demand Prediction

Sumin Han¹, Jisun An², Youngjun Park¹, Suji Kim³, Kitae Jang³, and Dongman Lee¹

¹ School of Computing, Korea Advanced Institute of Science and Technology (KAIST)

{hsm6911, youngjourpark, dlee}@kaist.ac.kr

² Luddy School of Informatics, Computing, and, Engineering Indiana University Bloomington

jisunan@iu.edu

³ The Cho Chun Shik Graduate School of Mobility, Korea Advanced Institute of Science and Technology (KAIST)

{sujikim, kitaejang}@kaist.ac.kr

Abstract. A reliable short-term transportation demand prediction supports the authorities in improving the capability of systems by optimizing schedules, adjusting fleet sizes, and generating new transit networks. A handful of research efforts incorporate one or a few areal features while learning spatio-temporal correlation, to capture similar demand patterns between similar areas. However, urban characteristics are polymorphic, and they need to be understood by multiple areal features such as land use, sociodemographics, and place-of-interest (POI) distribution. In this paper, we propose a novel spatio-temporal multi-feature-aware graph convolutional recurrent network (ST-MFGCRN) that fuses multiple areal features during spatio-temporal understanding. Inside ST-MFGCRN, we devise sentinel attention to calculate the areal similarity matrix by allowing each area to take partial attention if the feature is not useful. We evaluate the proposed model on two real-world transportation datasets, one with our constructed BusDJ dataset and one with benchmark TaxiBJ. Results show that our model outperforms the state-of-the-art baselines up to 7% on BusDJ and 8% on TaxiBJ dataset.

Keywords: Transportation Demand · Areal Feature · Multi-graph Convolutional RNN

1 Introduction

A reliable short-term transportation demand prediction supports the authorities in improving the capability of transportation systems. However, this task is challenging because the transit demand prediction depends on the complex spatial and temporal correlation [1]. With the development of deep learning approaches to short-term transportation demand prediction, several models have been proposed to deal with the spatio-temporal complexity in transportation

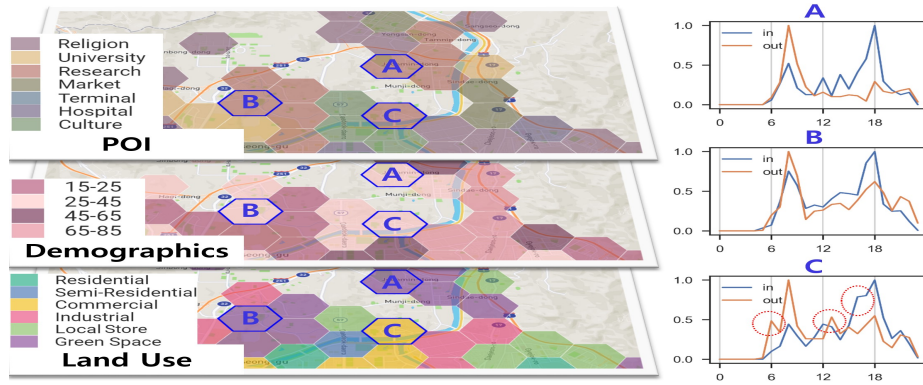


Fig. 1: Similar travel patterns of areas with the same POI and demographics and another data layer of land use (color represents the majority group).

demand [2,3,4,5,6]. For better prediction, first, temporal dependency should be considered as each area shows similar transportation demand according to similar time zones on the same day of the week, as well as historical transportation demand for each timestep [7,8]. Secondly, spatial dependency should also be incorporated. According to Tobler’s first law of geography [9], the travel demand of an area has significant spatial relation to the adjacent areas. Moreover, even distant areas with a similar urban environment (e.g., land use, point-of-interests (POIs)) show similar travel patterns due to the comparable purpose of visit [10,11,12].

A handful of deep learning models incorporate areal features while learning spatio-temporal correlation for transportation demand prediction, such as POI distribution [5,13,14,15] and demographics [6]. Despite promising results, we argue that one important aspect has been overlooked in those efforts; they represented an area with a very few features. Urban characteristics are polymorphic [16], and an area should be defined considering physical, social, and cultural aspects through multiple data layers such as land use, sociodemographics, building types, POI distribution, and transportation infrastructures [17,18]. For example, three areas (A, B, C) in Fig.1 show a similar travel pattern (more alighting in the morning and more boarding in the evening). This is expected as they all have many offices, and people aged 25-45 are commuting to those areas during workdays. However, there exists a difference observed in the area C – small peaks in the early morning (6h) and daytime (13h, 16h), which could be driven by other areal urban features such as land use. Thus, to better predict future travel demand, it is necessary to devise a new method that can leverage multiple areal features by considering their significance and correlation.

In this paper, we propose a spatio-temporal multi-feature-aware graph convolutional recurrent network (ST-MFGCRN) model that ensembles multiple areal similarity knowledge while preventing overfitting by sentinel attention⁴. First, to capture the spatial relatedness of different time intervals, we split time-series in-

⁴ The code, data, and supplemental materials are available in this anonymous repository: <https://anonymous.4open.science/r/ST-MFGCRN-251B/>

put data into three kinds of time units (closeness, period, and trend). Our model ingests each time-unit input with spatio-temporal embedding to capture trivial spatio-temporal patterns. Then, we use our multi-feature graph convolutional recurrent neural network to understand multiple areal similarity graphs while finding spatio-temporal correlations. Here, the areal similarity is computed by an attention-based mechanism from areal features. This attention mechanism is improved with sentinel variables that allow each area to take partial attention when the areal feature is not helpful. Finally, we merge the three time-unit outputs by our proposed weighted fusion mechanism to mix them properly. We evaluate the ST-MFGCRN model with other existing schemes against two real-world datasets. Here we summarize our contributions:

- We propose a novel deep-learning architecture for short-term transportation demand prediction that incorporates multiple areal features while learning spatio-temporal correlation.
- We analyze the impact of areal features to extract the best feature combination and highlight which areal features drive the best performance.
- Evaluation results show that the proposed ST-MFGCRN outperforms state-of-the-art base-lines such as DeepSTN+ and GMAN up to 7-8%, and is also resilient when many areal features are computed together.

2 Related Work

2.1 Spatio-temporal prediction methodology

For Euclidian dataset, a convolutional neural network (CNN) based spatio-temporal model is mainly used [2]. ConvLSTM [3] is a model modified for time series data by making 2D convolution in the matrix multiplication part of the weight parameter of a LSTM. STResNet [4] and DeepSTN+ [5] learn spatial information with 2D CNN for closeness, period, and trend, respectively, and time series temporal information with 1D CNN. DeepSTN+ is a slightly advanced form compared to STResNet in that it leverages a ConvPlus network, but both are similar as they are CNN-based models.

For non-Euclidian dataset, a graph convolutional network (GCN) based spatio-temporal model is mainly used [19,15]. DCRNN [20] proposes an RNN model through random walk-based diffusion convolution based on the distance of traffic sensor to learn graph-based spatio-temporal information. ASTGCN [21] leverages spatial and temporal attentional graph convolution mechanism on closeness, period, and trend components for road traffic prediction. GMAN [22] creates performs spatial attention with node embedding, and temporal attention, and combine with gated fusion.

2.2 Transit prediction with extra knowledge

A handful of researchers [5,4,13,6,15,23] have attempted to increase the prediction accuracy by using ancillary information about the area, such as POI

or demographics. There exist two major methods to utilize areal features—concatenating each areal feature along with input data or generating a similarity graph between areas and applying graph convolution. The former is a method of obtaining a specific pattern from the transit demand value itself, and the latter differs in that it utilizes information from areas with similar characteristics to inference when performing prediction for each area. [5,4,13] correspond to the former, although there are modifications such as using more fully connected layers or conducting addition instead of concatenation. [6,15,23,14] utilize Pearson correlation, cosine similarity, or community detection by the Louvain algorithm to construct similarity matrices.

2.3 Multi-graph spatio-temporal methodology

There are a few approaches that merges multiple graphs for spatio-temporal computation. DMVST-Net[24] suggests semantic view on top of spatial view and temporal view by creating data-driven semantic similarity graph by Dynamic Time Wrapping for taxi demand prediction. ST-MGCN[14] proposes multi-graph convolution network consists of neighborhood, functional similarity, and spatial connectivity for ride-hailing demand forecasting. Wang et al.[25] proposes heterogeneous multi-graph convolution network consists of spatial adjacency, transport connectivity, and contextual similarity for ambulance demand forecasting. AGCAN[26] leverages data-driven attention-based adaptive graph to supplement along with physical connectivity graph for road traffic prediction.

3 Problem formulation

We define that the number of transportation areas is N , the number of input channels is C_x (e.g. $C_x = 2$ for in/out demand of each area), and the demand (e.g., number of passengers who aboard/depart the bus) on each timestamp at $X_t \in \mathbb{R}^{N \times C_x}$, and the spatial proximity graph between areas is \mathcal{G}^P . Assume we want to use K_N types of areal features, where k -th feature is $F_k \in \mathbb{R}^{N \times V_k}$, where V_k is the number of components of F_k (e.g. $V_{LU} = 3$ when the land use feature F_{LU} consists of commercial/residential/office). The transportation demand prediction problem can be defined as learning a model f_{model} that predicts the next-step value from historical values as:

$$f_{\text{model}}(X_1, X_2, \dots, X_{t-1}; \mathcal{G}^P, F_{\{1, \dots, K_N\}}) \rightarrow X_t \quad (1)$$

4 Method

4.1 Overview

We propose spatio-temporal multi-feature graph convolutional recurrent network (ST-MFGCRN) described in Fig. 2a. Our method splits the historical value into

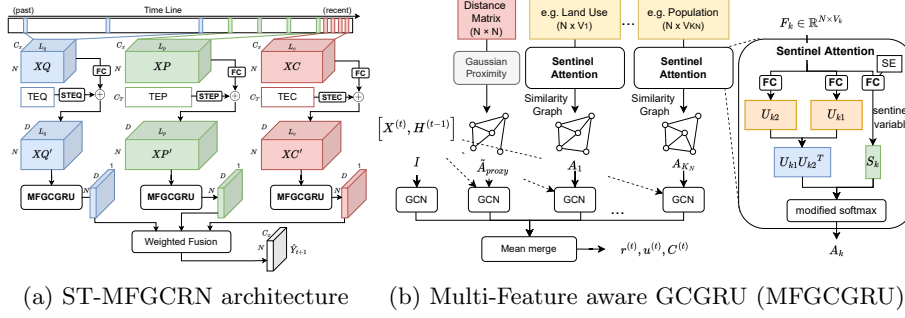


Fig. 2: (a) Proposed ST-MFGCRN architecture and (b) MFGCGRU cell. TEQ, TEP, TEC represents the temporal embedding (TE) for trend, period, and closeness modules which convert into spatio-temporal embedding (STE).

three different time units – closeness (recent minutes), period (daily), and trend (weekly) similar to previous approaches[5,4,27] as follows:

$$\begin{aligned}
 \mathbf{XC} &= \{X_t, \dots, X_{t-L_c+1}\} \in \mathbb{R}^{N \times L_c \times C_x}, \\
 \mathbf{XP} &= \{X_{t-T_p+1}, \dots, X_{t-T_p L_p+1}\} \in \mathbb{R}^{N \times L_p \times C_x}, \\
 \mathbf{XQ} &= \{X_{t-T_q+1}, \dots, X_{t-T_q L_q+1}\} \in \mathbb{R}^{N \times L_q \times C_x}
 \end{aligned} \tag{2}$$

where L_c , L_p , and L_q are the number of timesteps of closeness, period, and trend values, respectively. T_c is the number of close sequences, and T_p is the number of steps between periods while T_q is that for trends. For hidden input and output between all modules, we use D as the embedding dimension in common.

4.2 Spatio temporal embedding (STE)

Spatio temporal embedding (STE) serves to capture and correct trivial spatio-temporal bias for each area. Compared to GMAN [22] that leverages STE that serves similar to positional encoding, our model directly adds to embedded input value. We preprocess the temporal embedding (TE) as $\text{TE}_t \in \mathbb{R}^{C_T}$ by concatenating each one-hot embedding of a time unit, where $C_T = 7 + 24 + 4 + 1$: weekday (7), hour-of-day (24), 15-min-unit-of-hour (4), national holiday (1). Then, we apply a two-stacked fully connected layer $f_{\text{TE}} : \mathbb{R}^{C_T} \rightarrow \mathbb{R}^D$ to convert TE into D -dimensional hidden embedding. Moreover, we define a trainable spatial embedding $\text{SE} \in \mathbb{R}^{N \times D}$ and make the spatio-temporal embedding as $\text{STE}_t = \text{SE} + f_{\text{TE}}(\text{TE}_t) \in \mathbb{R}^{N \times D}$. Meanwhile, we extend the initial traffic input $X_t \in \mathbb{R}^{N \times C_x}$ into $f_{\text{in}}(X_t) \in \mathbb{R}^{N \times D}$ by applying a two-stacked fully connected layer $f_{\text{in}} : \mathbb{R}^{C_x} \rightarrow \mathbb{R}^D$, and produce the traffic value embedding normalized with STE as $X'_t = f_{\text{in}}(X_t) + \text{STE}_t \in \mathbb{R}^{N \times D}$ and use it as a next input of a module. Here, we share the same SE and f_{TE} for different time units, while $f_{\text{in}}^{(c)}$, $f_{\text{in}}^{(p)}$, $f_{\text{in}}^{(q)}$ differently used for each \mathbf{XC} , \mathbf{XP} , \mathbf{XQ} .

4.3 Multi-feature aware GCGRU (MFGCGRU)

We propose a multi-feature aware GCGRU (MFGCGRU) by extending graph convolutional gated recurrent unit to compute multiple areal features inside GRU cell. At first, the model computes spatial proximity $\mathcal{G}^P(A_{prox})$ by a graph convolution. A proximity matrix is calculated from distances with a Gaussian filter as $A_{prox}[i, j] = \exp(-(d_{ij}/\sigma)^2)$, where the d_{ij} is the distance between i -th and j -th areas and σ is the standard deviation of distance values. After we conduct row-normalization as \tilde{A}_{prox} , we use this matrix as one of graphs computed in MFGCGRU.

Next, in order to extract areal similarity matrix $A_k \in \mathbb{R}^{N \times N}$ from areal feature F_k to be used in MFGCGRU, we leverage attention mechanism as follows:

$$\begin{aligned} U_{k1} &= \text{ReLU}(\mathbf{W}_{k1}F_k) \in \mathbb{R}^{N \times D}, & U_{k2} &= \text{ReLU}(\mathbf{W}_{k2}F_k) \in \mathbb{R}^{N \times D}, \\ e_k &= U_{k1}U_{k2}^T/\sqrt{D} \in \mathbb{R}^{N \times N}, & A_k[i, j] &= \frac{\exp(e_k[i, j])}{\sum_o \exp(e_k[i, o])} \end{aligned} \quad (3)$$

, where $\mathbf{W}_{k1}, \mathbf{W}_{k2} \in \mathbb{R}^{D \times V_k}$ are trainable parameters for each feature type k . e_k is basically represents the similarity between U_{k1} and U_{k2} by applying dot product for each $N \times N$ area pairs.

However, as more areal features are added, the model is congested and do not reflect the areal similarity due to the curse of dimensionality[28]. Inspired by a sentinel attention [29,30], we calculate the attention score for each area can be less than 1 by applying sentinel variable S_k , which allows an area to take partial attention when the areal knowledge is not beneficial. Our proposed sentinel attention is defined as follows:

$$\begin{aligned} S_k &= f_{\text{sent}}(F_k \parallel \text{SE}) \in \mathbb{R}^N, \\ A_k[i, j] &= \frac{\exp(e_k[i, j])}{S_k[i] + \sum_o \exp(e_k[i, o])} \end{aligned} \quad (4)$$

, where f_{sent} is a two-stacked fully connected layer with ReLU activation (which keeps $S_k \geq 0$). We apply concatenation ($F_k \parallel \text{SE}$) to calculate sentinel variable differently for each area using spatial embedding.

$$\begin{aligned} \mathbf{r}^{(t)} &= \sigma \left(\frac{1}{K} \sum_{k=1}^K A_k \left(\mathbf{X}^{(t)} \parallel \mathbf{H}^{(t-1)} \right) \mathbf{W}_{rk} + \mathbf{b}_r \right) \\ \mathbf{u}^{(t)} &= \sigma \left(\frac{1}{K} \sum_{k=1}^K A_k \left(\mathbf{X}^{(t)} \parallel \mathbf{H}^{(t-1)} \right) \mathbf{W}_{uk} + \mathbf{b}_u \right) \\ \mathbf{C}^{(t)} &= \tanh \left(\frac{1}{K} \sum_{k=1}^K A_k \left(\mathbf{X}^{(t)} \parallel \left(\mathbf{r}^{(t)} \odot \mathbf{H}^{(t-1)} \right) \right) \mathbf{W}_{Ck} + \mathbf{b}_c \right) \\ \mathbf{H}^{(t)} &= \mathbf{u}^{(t)} \odot \mathbf{H}^{(t-1)} + \left(1 - \mathbf{u}^{(t)} \right) \odot \mathbf{C}^{(t)}. \end{aligned} \quad (5)$$

Finally, we ensemble multiple graph convolution inside GRU cell by applying mean merge of multiple graphs as Equation 5, where $\mathbf{W}_{rk}, \mathbf{W}_{uk}, \mathbf{W}_{Ck} \in \mathbb{R}^{(D_x+D_h) \times D_h}$ (for $k = 1, \dots, K$) and $\mathbf{b}_r, \mathbf{b}_u, \mathbf{b}_C \in \mathbb{R}^{D_h}$ are trainable parameters, and \odot is Hadamard product. We also apply an identity matrix (I) on top of a proximity matrix and K_N area similarity matrices as a residual computation, thus $K = 2 + K_N$ as a default.

4.4 Weighted fusion network

From each MFGCGRU for $\mathbf{XC}, \mathbf{XP}, \mathbf{XQ}$, we get an output of $H_C, H_P, H_Q \in \mathbb{R}^{N \times D}$, respectively. We apply a fully connected layer and a softmax to calculate the weights of importance, and produce final output with a two-stacked fully connected layer $f_{\text{fusion}} : \mathbb{R}^D \rightarrow \mathbb{R}^{C_x}$ as follows, where $\mathbf{W}_C, \mathbf{W}_P, \mathbf{W}_Q \in \mathbb{R}^{1 \times D}$ are trainables:

$$e_m = \mathbf{W}_m H_m \in \mathbb{R}^N (m \in \{C, P, Q\}), \quad \alpha_m = \frac{\exp(e_m)}{\sum_j^{\{C, P, Q\}} \exp(e_j)} \in \mathbb{R}^N, \quad (6)$$

$$\hat{Y}_{t+1} = f_{\text{fusion}} \left(\sum_j^{\{C, P, Q\}} \alpha_j \odot H_j \right) \in \mathbb{R}^{N \times C_x}.$$

4.5 Objective function

The objective function is defined as $\mathcal{L}(\theta) = \frac{1}{N} \sum_i^N |Y_{t+1}^i - \hat{Y}_{t+1}^i|$ where $Y_t \in \mathbb{R}^{N \times C_x}$, which is the *mean absolute error* (L1) loss.

5 Evaluation settings

5.1 Dataset

We use two datasets in our experiments: our constructed BusDJ dataset and a benchmark TaxiBJ dataset. For the BusDJ dataset, we use the bus transit data of boarding and alighting demand in a 15-minute unit which we obtain from smart card data of Daejeon Metropolitan City. We construct the bus transit dataset of $N = 102$ hexagonal 800m grid areas except the areas containing no bus stop (e.g., hills, rivers). Then, we leverage six areal features representing land use, transit, and demographic features: (1) land use type (LU, $V_{LU} = 6$), (2) place-of-interest (POI, $V_{POI} = 29$), (3), bus transit infrastructure (BUS, $V_{BUS} = 2$), (4) population (POP, $V_{POP} = 4$), (5) enterprise and employments of each business type (ENT, $V_{ENT} = 6$), and (6) building area (BD, $V_{BD} = 7$). The areal features are provided by the open data service of government institutions [31,32,33]. For benchmark TaxiBJ [13] dataset, we use two available areal features provided in the original dataset. We extract the 20 most frequent POI features among the original 648 types of POIs ($V_{POI} = 20$). We also use two ROAD features (the number of roads and lanes) as original data ($V_{ROAD} = 2$). The detailed description of both data is listed in Table 1.

Table 1: Details of the datasets.

Dataset	BusDJ	TaxiBJ
N (area size)	102 (800m Hexa)	146 (1km Square)
Train timespan	3/1/2019 - 4/30/2019	2/1/2015 - 6/2/2015
Val timespan	5/1/2019 - 5/31/2019	6/3/2015 - 6/16/2015
Test timespan	6/1/2019 - 6/30/2019	6/17/2015 - 6/30/2015
Time interval	15min	30min
# timestamps	6832	7152
Average demand (in/out)	32.05/31.41	626.10/625.35
Target prediction time	8h - 21h (14h)	0h - 23h (24h)
# areal features	6	2

5.2 Experiment settings

We set the number of time sequences as $L_c = 6, L_p = 7, L_q = 3$ in Eq. 2, and $C_x = 2$ for boarding and alighting demand. In training, we conduct min-max normalization on demand and areal features. We set the hidden embedded dimension for our model as $D = 64$. For other baselines, we use their default settings. We conduct our experiments on a single NVIDIA TITAN RTX 24GB environment with Tensorflow v1.15.1. For each model on each setting, we test five times and record the mean of each result. We adopt an ADAM optimizer with a learning rate of 0.01 and a batch size of 32. We apply early stopping with validation loss with 15 patience epochs. We evaluate models with *root mean squared error* (RMSE) and *mean absolute error* (MAE). We measure the error of C_x input channels altogether.

5.3 Baselines

We compare the proposed model (ST-MFGCRN) with various methods for public transit prediction. We first use data-based prediction methods without model training as basic baselines: Trend Mean, Period Mean, Closeness Mean, and Last Repeat. We set CNN-based time-series deep learning models, such as ConvLSTM [3], STResNet [4], and DeepSTN+ [5]. We also compare our model with the graph-based deep learning models such as DCRNN [20], ASTGCN [21], STMGCN [14], and GMAN [22] by using proximity information A_{prox} described in Section 4.3. For STMGCN, we use cosine-similarity based adjacency matrix for each feature for graph computation. For ASTGCN and STMGCN, we use closeness, period, trend component same as our model. For other baselines, we use the default settings.

6 Results

We first analyze the feature influence by adding or removing each areal feature from our model to extract the most important areal feature combination. Then, we evaluate the performance of ST-MFGCRN by comparing it with baseline models on the BusDJ and TaxiBJ datasets.

6.1 Analysis of feature significance

Table 2: Feature influence on BusDJ (RMSE). Results of important features are underlined.

	Feature add/remove	$-\mathcal{G}^P$	$+\mathcal{G}^P$
$K_N = 0$	N/A		8.5032
$K_N = 1$	+LU	8.5000	<u>8.4625</u>
	+POI	8.5409	<u>8.4675</u>
	+BUS	8.5141	8.4747
	+POP	8.5063	8.5119
	+ENT	8.5100	<u>8.4638</u>
	+BD	8.5290	8.5031
$K_N = 5$	(all)–LU	8.4941	8.4635
	(all)–POI	8.4939	<u>8.4784</u>
	(all)–BUS	8.4976	8.4670
	(all)–POP	8.4976	8.4619
	(all)–ENT	8.4937	<u>8.4816</u>
	(all)–BD	8.4782	<u>8.4963</u>
$K_N = 6$	(all)	8.5005	8.4606
Best	+BD, +ENT, +POI, +LU		<u>8.4587</u>

Table 3: Feature influence on TaxiBJ dataset.

\mathcal{G}^P	F_{POI}	F_{ROAD}	RMSE
–	–	–	99.849
+	–	–	95.487
+	+	–	93.860
+	–	+	92.782
+	+	+	<u>92.767</u>

To examine to what extent individual features improve prediction performance, we conduct a feature influence study by adding or removing features.

For the BusDJ dataset, Table 2 shows a significant error decrease when the proximity graph \mathcal{G}^P is provided, implying that geographical closeness essentially captures the basic spatial correlation and improves the performance. The experiment results by all combinations (when $K_N = 2, 3, 4, 5$) are omitted due to the space limit. We find that LU, ENT, and POI are the most significant areal features when added, as their errors decrease the most from $\{K_N = 0, +\mathcal{G}^P\}$ model. On the other hand, when we evaluate the most significant areal features when removed, we observe BD, ENT, and POI increase the most errors from $\{K_N = 6, +\mathcal{G}^P\}$ model. Throughout all experiments, the best setting is found as a combination of BD, ENT, POI, and LU.

For TaxiBJ dataset, Table 3 shows the POI and ROAD features do improve the performance significantly. The model shows the best result when it takes all the areal features as well as the proximity graph.

These results imply that each areal feature provides unique information about the urban environment, and multiple views of urban area supplements the overall performance.

Table 4: Performance comparison.

Model	BusDJ		TaxiBJ	
	RMSE	MAE	RMSE	MAE
Trend Mean	11.3647	6.8445	307.430	199.915
Period Mean	15.3843	8.8179	309.706	198.528
Closeness Mean	15.8035	8.9490	225.092	159.634
Last Repeat	13.6378	8.6482	116.877	72.842
ConvLSTM	9.2843	5.9005	100.939	57.500
STResNet	9.2875	5.9241	103.689	63.305
DeepSTN+	9.2215	5.8746	107.319	64.830
DCRNN	9.5296	6.1572	107.138	66.566
ASTGCN	10.7053	6.5875	126.197	75.555
STMGCN	10.0883	6.3992	110.993	71.346
GMAN	9.1381	5.8662	102.276	56.109
DCRNN+STE+CPT	8.7775	5.6461	106.756	62.876
GMAN+STE+CPT	8.9188	5.7136	113.828	67.689
ST-MFGCRN (best*)	8.4635	5.4987	92.767	52.443

*BusDJ: BD, ENT, POI, LU, TaxiBJ: POI, ROAD

6.2 Performance comparison

Table 4 shows the performance comparison of models on BusDJ and TaxiBJ datasets. Our proposed ST-MFGCRN shows an improved performance of 7% on BusDJ dataset and 8% on the TaxiBJ dataset compared to the state-of-the-art baselines such as DeepSTN+ and GMAN in terms of RMSE. Among the basic baselines, Trend Mean shows the best performance, which implies that the travel patterns are similar for the same hour and weekday every week. On the other hand, Closeness Mean and Last Repeat show high errors, indicating the need for a temporal correlation training model. Among the Euclidian area-based models, DeepSTN+ shows the highest performance as DeepSTN+ is more advanced than STResnet or ConvLSTM. In addition, ConvLSTM shows better performance than STResNet which leverages LSTM instead of 1D-CNN to find the temporal correlation. Among the non-Euclidian models, GMAN shows the highest performance compared to DCRNN and ASTGCN as it leverages spatial attention which allows attention on all other areas rather than graph convolution with fixed proximity values.

We also conduct experiments when we expand DCRNN with STE and close-ness/period/trend (CPT) unit and GMAN with CPT unit, and find each component improves performance on these baselines. However, our ST-MFGCRN still outperforms as it leverages multiple features with sentinel attention and applies weighted fusion to utilize these components.

7 Conclusion

We propose ST-MFGCRN which allows exploiting various areal features during spatio-temporal training to tackle transportation demand prediction problems. To capture correlations depending on areal characteristics and spatial adjacency, ST-MFGCRN incorporates multiple similarity matrices calculated from different areal features and a distance-based proximity matrix. To calculate the similarity matrix of each areal feature, we leverage our proposed sentinel attention, which plays a role in taking partial attention when the knowledge is not helpful and harmoniously mixes various features. ST-MFGCRN captures temporal correlation by different time unit inputs of closeness, period, and trend, and spatio-temporal dependencies are also captured by spatio-temporal embedding. Our model outperforms the state-of-the-art baselines on the real-world dataset that we construct as well as on the benchmark dataset while successfully leveraging multiple areal features.

References

1. C. Morency, M. Trépanier, and B. Agard, “Measuring transit use variability with smart-card data,” *Transport Policy*, vol. 14, no. 3, pp. 193–203, 2007.
2. M. Luca, G. Barlacchi, B. Lepri, and L. Pappalardo, “A survey on deep learning for human mobility,” *ACM Computing Surveys (CSUR)*, vol. 55, no. 1, pp. 1–44, 2021.
3. X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, “Convolutional lstm network: A machine learning approach for precipitation nowcasting,” *Advances in neural information processing systems*, vol. 28, 2015.
4. J. Zhang, Y. Zheng, and D. Qi, “Deep spatio-temporal residual networks for city-wide crowd flows prediction,” in *Thirty-first AAAI conference on artificial intelligence*, 2017.
5. Z. Lin, J. Feng, Z. Lu, Y. Li, and D. Jin, “Deepstn+: Context-aware spatial-temporal neural network for crowd flow prediction in metropolis,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 1020–1027, 2019.
6. H. Liu, Q. Wu, F. Zhuang, X. Lu, D. Dou, and H. Xiong, “Community-aware multi-task transportation demand prediction,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 320–327, 2021.
7. S. Foell, S. Phithakkitnukoon, M. Veloso, G. Kortuem, and C. Bento, “Regularity of public transport usage: A case study of bus rides in lisbon, portugal,” *Journal of Public Transportation*, vol. 19, no. 4, p. 10, 2016.
8. X. Ma, Y.-J. Wu, Y. Wang, F. Chen, and J. Liu, “Mining smart card data for transit riders’ travel patterns,” *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 1–12, 2013.
9. W. R. Tobler, “A computer movie simulating urban growth in the detroit region,” *Economic geography*, vol. 46, no. sup1, pp. 234–240, 1970.
10. C. Chen, J. Chen, and J. Barry, “Diurnal pattern of transit ridership: a case study of the new york city subway system,” *Journal of Transport Geography*, vol. 17, no. 3, pp. 176–186, 2009.
11. Z. Gan, M. Yang, T. Feng, and H. Timmermans, “Understanding urban mobility patterns from a spatiotemporal perspective: daily ridership profiles of metro stations,” *Transportation*, vol. 47, no. 1, pp. 315–336, 2020.

12. S. Ling, Z. Yu, S. Cao, H. Zhang, and S. Hu, "Sthan: Transportation demand forecasting with compound spatio-temporal relationships," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2022.
13. Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, and J. Zhang, "Urban traffic prediction from spatio-temporal data using deep meta learning," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1720–1730, 2019.
14. X. Geng, Y. Li, L. Wang, L. Zhang, Q. Yang, J. Ye, and Y. Liu, "Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 3656–3663, 2019.
15. J. Tang, J. Liang, F. Liu, J. Hao, and Y. Wang, "Multi-community passenger demand prediction at region level based on spatio-temporal graph convolutional network," *Transportation Research Part C: Emerging Technologies*, vol. 124, p. 102951, 2021.
16. S. Harrison and P. Dourish, "Re-place-ing space: the roles of place and space in collaborative systems," in *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pp. 67–76, 1996.
17. B. D. Taylor, D. Miller, H. Iseki, and C. Fink, "Nature and/or nurture? analyzing the determinants of transit ridership across us urbanized areas," *Transportation Research Part A: Policy and Practice*, vol. 43, no. 1, pp. 60–77, 2009.
18. R. Ewing and R. Cervero, "Travel and the built environment: A meta-analysis," *Journal of the American planning association*, vol. 76, no. 3, pp. 265–294, 2010.
19. W. Jiang and J. Luo, "Graph neural network for traffic forecasting: A survey. arxiv," *arXiv preprint arXiv:2101.11174*, 2021.
20. Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," *arXiv preprint arXiv:1707.01926*, 2017.
21. S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 922–929, 2019.
22. C. Zheng, X. Fan, C. Wang, and J. Qi, "Gman: A graph multi-attention network for traffic prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 1234–1241, 2020.
23. H. Yuan, G. Li, Z. Bao, and L. Feng, "An effective joint prediction model for travel demands and traffic flows," in *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pp. 348–359, IEEE, 2021.
24. H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, and Z. Li, "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018.
25. Z. Wang, T. Xia, R. Jiang, X. Liu, K.-S. Kim, X. Song, and R. Shibasaki, "Forecasting ambulance demand with profiled human mobility via heterogeneous multi-graph neural networks," in *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pp. 1751–1762, IEEE, 2021.
26. B. Li, T. Guo, Y. Wang, A. H. Gandomi, and F. Chen, "Adaptive graph co-attention networks for traffic forecasting," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 263–276, Springer, 2021.
27. J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi, "Dnn-based prediction model for spatio-temporal data," in *Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems*, pp. 1–4, 2016.
28. P. Indyk and R. Motwani, "Approximate nearest neighbors: towards removing the curse of dimensionality," in *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pp. 604–613, 1998.

29. J. Lu, C. Xiong, D. Parikh, and R. Socher, “Knowing when to look: Adaptive attention via a visual sentinel for image captioning,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 375–383, 2017.
30. C. Park, C. Lee, H. Bahng, Y. Tae, S. Jin, K. Kim, S. Ko, and J. Choo, “St-grat: A novel spatio-temporal graph attention networks for accurately forecasting dynamically changing road speed,” in *Proceedings of the 29th ACM International conference on information & knowledge management*, pp. 1215–1224, 2020.
31. M. of the Interior and Safety, “Local data.”
32. I. Ministry of Land and Transport, “Korea national spatial infrastructure portal.”
33. S. Korea, “Statistical geographic information service.”