Time Series

王慧琪

200906

KDD2020

Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks

Zonghan Wu
University of Technology Sydney
zonghan.wu-3@student.uts.edu.au

Shirui Pan*
Monash University
shirui.pan@monash.edu

Guodong Long
University of Technology Sydney
guodong.long@uts.edu.au

Jing Jiang
University of Technology Sydney
jing.jiang@uts.edu.au

Xiaojun Chang
Monash University
xiaojun.chang@monash.edu

Chengqi Zhang
University of Technology Sydney
chengqi.zhang@uts.edu.au

Motivation

- *Challenge 1: Unknown Graph Structure.* Existing GNN approaches rely heavily on a pre-defined graph structure in order to perform time series forecasting. In most cases, multivariate time series does not have an explicit graph structure. The relationships among variables has to be discovered from data rather than being provided as ground truth knowledge.
- *Challenge 2: Graph Learning & GNN Learning.* Even though a graph structure is available, most GNN approaches focus only on message passing (GNN Learning) and overlook the fact that the graph structure is not optimal and should be updated during training. The question then is how to simultaneously learn the graph structure and the GNN for time series in an end-to-end framework.

PROBLEM FORMULATION

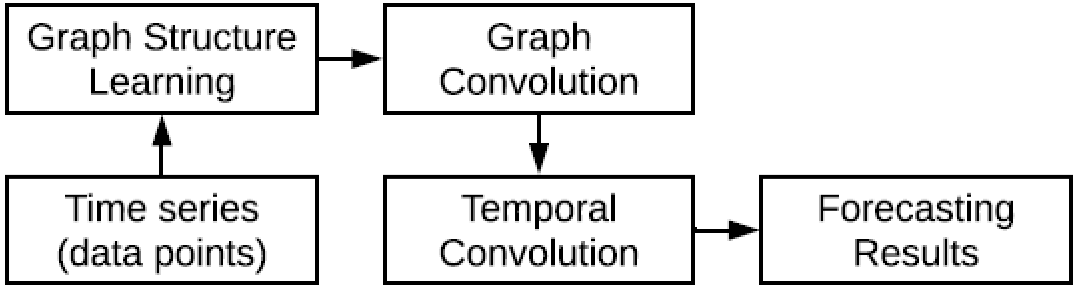
In this paper, we focus on the task of multivariate time series forecasting. Let $\mathbf{z}_t \in \mathbf{R}^N$ denote the value of a multivariate variable of dimension N at time step t, where $z_t[i] \in R$ denote the value of the i^{th} variable at time step t. Given a sequence of historical P time steps of observations on a multivariate variable, $\mathbf{X} = \{\mathbf{z}_{t_1}, \mathbf{z}_{t_2}, \cdots, \mathbf{z}_{t_P}\}$, our goal is to predict the Q-step-away value of $\mathbf{Y} = \{\mathbf{z}_{t_{P+Q}}\}$, or a sequence of future values $\mathbf{Y} = \{\mathbf{z}_{t_{P+1}}, \mathbf{z}_{t_{P+2}}, \cdots, \mathbf{z}_{t_{P+Q}}\}$. More generally, the input signals can be coupled with other auxiliary features such as time of the day, day of the week, and day of the season. Concatenating the input signals with auxiliary features, we assume the inputs instead are $\mathcal{X} = \{\mathbf{S}_{t_1}, \mathbf{S}_{t_2}, \cdots, \mathbf{S}_{t_P}\}$ where $\mathbf{S}_{t_i} \in \mathbf{R}^{N \times D}$, D is the feature dimension, the first column of $\mathbf{S}_{t_i}$ equals to $\mathbf{z}_{t_i}$, and the rest are auxiliary features. We aim to build a mapping $f(\cdot)$ from $\mathcal{X}$ to $\mathbf{Y}$ by minimizing the absolute loss with $l2$ regularization.

Definition 3.1 (Graph). A graph is formulated as $G = (V, E)$ where V is the set of nodes, and E is the set of edges. We use N to denote the number of nodes in a graph.

Definition 3.2 (Node Neighborhood). Let $v \in V$ to denote a node and $e = (v, u) \in E$ to denote an edge pointing from u to v. The neighborhood of a node v is defined as $N(v) = \{u \in V | (v, u) \in E\}$.

Definition 3.3 (Adjacency Matrix). The adjacency matrix is a mathematical representation of a graph, denoted as $\mathbf{A} \in R^{N \times N}$ with $A_{ij} = c > 0$ if $(v_i, v_j) \in E$ and $A_{ij} = 0$ if $(v_i, v_j) \notin E$.

Model Architecture



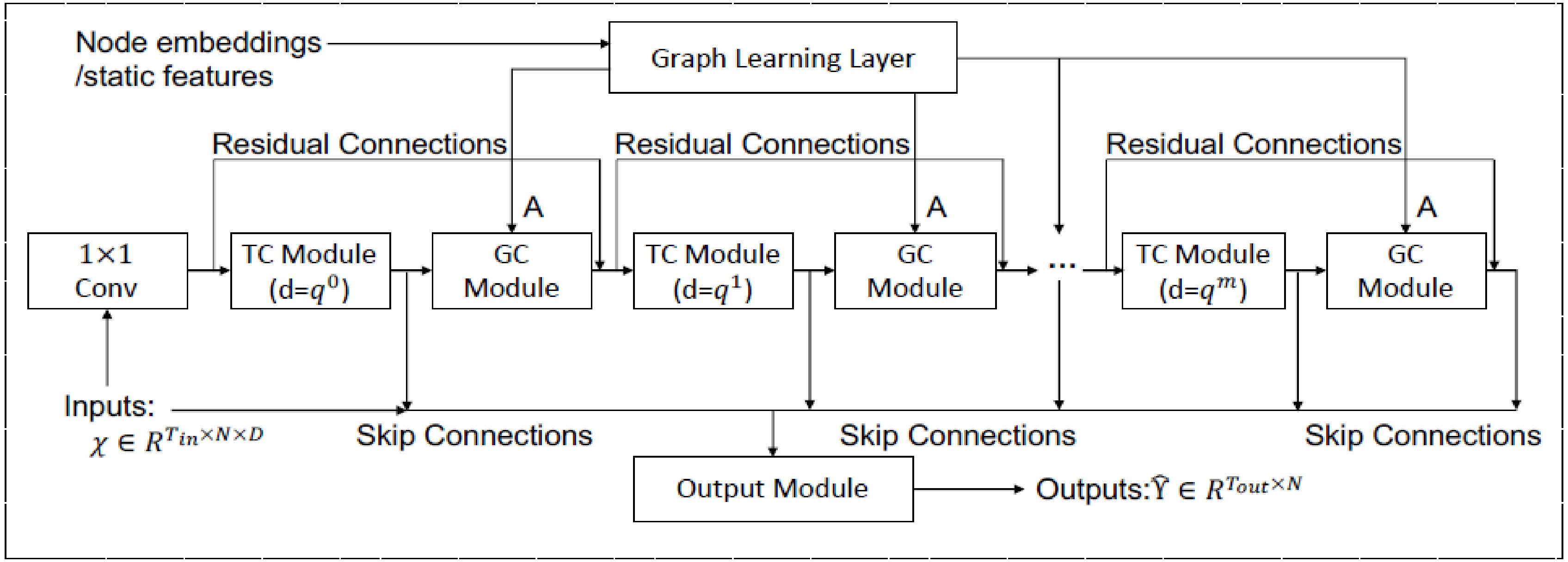


Figure 2: The framework of MTGNN. A 1×1 standard convolution first projects the inputs into a latent space. Afterward, temporal convolution modules and graph convolution modules are interleaved with each other to capture temporal and spatial dependencies respectively. The hyper-parameter, dilation factor d, which controls the receptive field size of a temporal convolution module, is increased at an exponential rate of q. The graph learning layer learns the hidden graph adjacency matrix, which is used by graph convolution modules. Residual connections and skip connections are added to the model to avoid the problem of gradient vanishing. The output module projects hidden features to the desired dimension to get the final results.

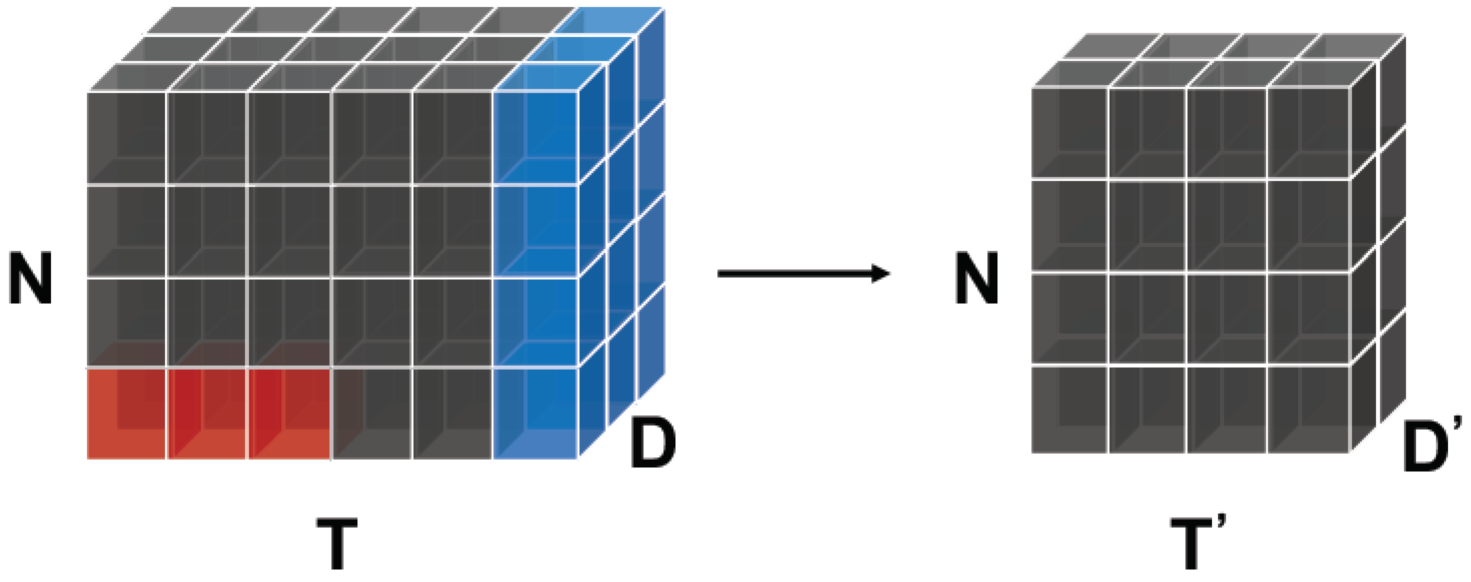
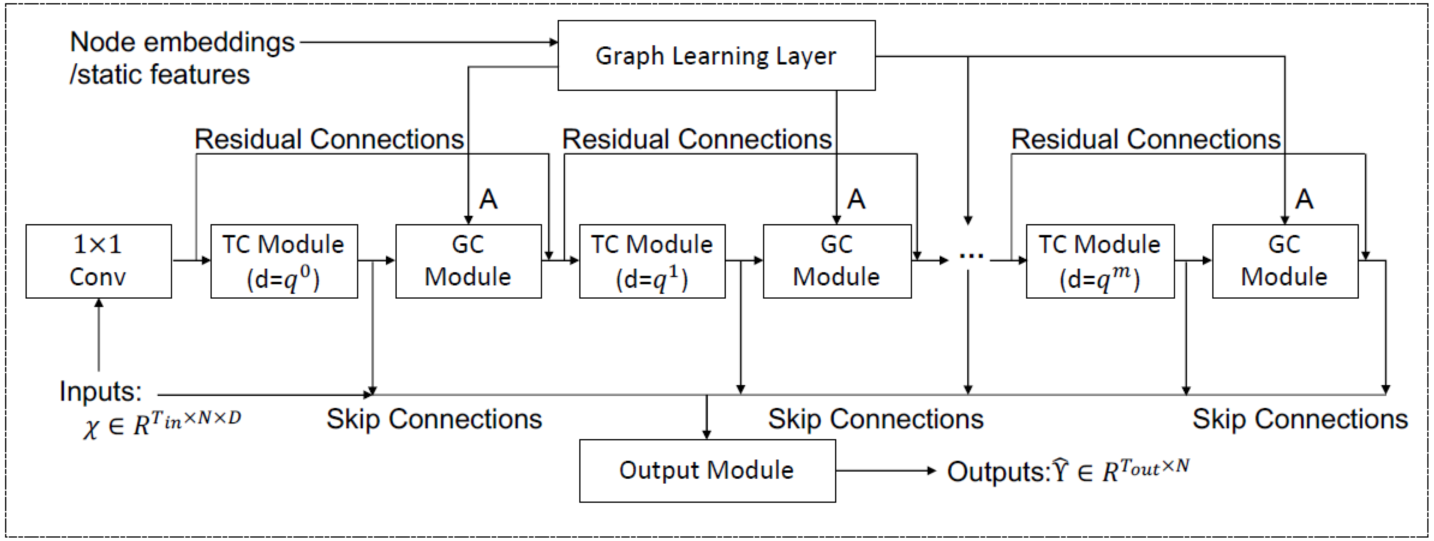


Figure 3: A demonstration of how a temporal convolution module and a graph convolution module collaborate with each other. A temporal convolution module filters the inputs by sliding a 1D window over the time and node axes, as denoted by the red. A graph convolution module filters the inputs at each step, denoted by the blue.

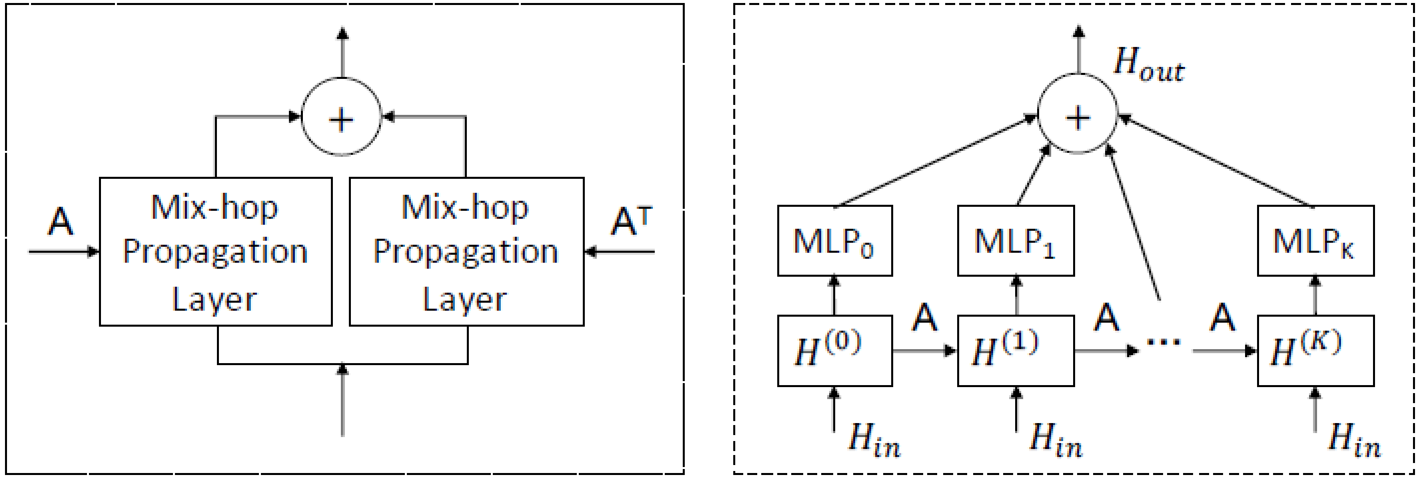
Graph Learning Layer



where $\mathbf{E}_1, \mathbf{E}_2$ represents randomly initialized node embeddings, which are learnable during training, Θ_1, Θ_2 are model parameters, α is a hyper-parameter for controlling the saturation rate of the activation function, and $argtopk(\cdot)$ returns the index of the top-k largest values of a vector. The asymmetric property of our proposed graph adjacency matrix is achieved by Equation 3. The subtraction term and the ReLU activation function regularize the adjacency matrix so that if A_{vu} is positive, its diagonal counterpart A_{uv} will be zero. Equation 5-6 is a strategy to make the adjacency matrix sparse while reducing the computation cost of the following graph convolution. For each node, we select its top-k closest nodes as its neighbors. While retaining the weights for connected nodes, we set the weights of non-connected nodes as zero.

$$\mathbf{M}_1 = tanh(\alpha \mathbf{E}_1 \Theta_1)$$
$$\mathbf{M}_2 = tanh(\alpha \mathbf{E}_2 \Theta_2)$$
$$\mathbf{A} = ReLU(tanh(\alpha(\mathbf{M}_1 \mathbf{M}_2^T - \mathbf{M}_2 \mathbf{M}_1^T)))$$
$$for\ i = 1, 2, \cdots, N$$
$$\mathbf{idx} = argtopk(\mathbf{A}[i, :])$$
$$\mathbf{A}[i, -\mathbf{idx}] = 0,$$

Graph Convolution Module



(a) GC module (b) Mix-hop propagation layer

Figure 4: Graph convolution and mix-hop propagation layer.

$$\mathbf{H}^{(k)} = \beta \mathbf{H}_{in} + (1-\beta)\tilde{\mathbf{A}}\mathbf{H}^{(k-1)},$$

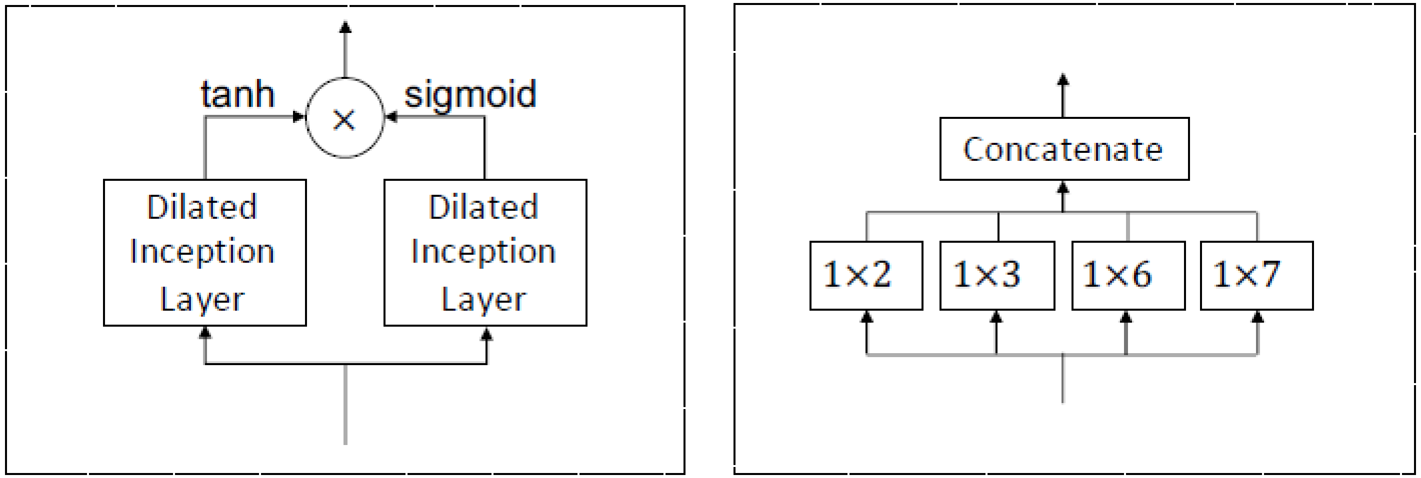
$$\mathbf{H}_{out} = \sum_{i=0}^{K} \mathbf{H}^{(k)}\mathbf{W}^{(k)},$$

Suppose $K = 2$, $\mathbf{W}^{(0)} = \mathbf{0}$, $\mathbf{W}^{(1)} = -\mathbf{1}$, and $\mathbf{W}^{(2)} = \mathbf{1}$, then

$$\mathrm{H}_{out} = \Delta(\mathrm{H}^{(2)}, \mathrm{H}^{(1)}) = \mathrm{H}^2 - \mathrm{H}^1. \tag{9}$$

where K is the depth of propagation, $\mathbf{H}_{in}$ represents the input hidden states outputted by the previous layer, $\mathbf{H}_{out}$ represents the output hidden states of the current layer, $\mathbf{H}^{(0)} = \mathbf{H}_{in}$, $\tilde{\mathbf{A}} = \tilde{\mathbf{D}}^{-1}(\mathbf{A}+\mathbf{I})$, and $\tilde{\mathbf{D}}_{ii} = 1 + \sum_j \mathbf{A}_{ij}$. In Figure 4b, we demonstrate the information propagation step and information selection step in the proposed mix-hop propagation layer. It first propagates information horizontally and selects information vertically.

Temporal Convolution Module



(a) TC module　　(b) Dilated inception layer

Figure 5: The temporal convolution and dilated inception layer.

$$R = 1 + (c-1)(q^m - 1)/(q-1).$$

$$\mathbf{z} = concat(\mathbf{z} \star \mathbf{f}_{1\times 2}, \mathbf{z} \star \mathbf{f}_{1\times 3}, \mathbf{z} \star \mathbf{f}_{1\times 6}, \mathbf{z} \star \mathbf{f}_{1\times 7}),$$

$$\mathbf{z} \star \mathbf{f}_{1\times k}(t) = \sum_{s=0}^{k-1} \mathbf{f}_{1\times k}(s)\mathbf{z}(t - d \times s),$$

Algorithm 1 The learning algorithm of MTGNN.

Input: The dataset O, node set V, the initialized MTGNN model $f(\cdot)$ with Θ, learning rate γ, batch size b, step size s, split size m (default=1).
set $iter = 1$, $r = 1$
repeat
    sample a batch ($\mathcal{X} \in R^{b \times T \times N \times D}$, $\mathcal{Y} \in R^{b \times T' \times N}$) from O.
    random split the node set V into m groups, $\cup_{i=1}^{m} V_i = V$.
    if $iter \% s == 0$ and $r <= T'$ **then**
        $r = r + 1$
    end if
    for i in 1:m **do**
        compute $\hat{\mathcal{Y}} = f(\mathcal{X}[:, :, id(V_i), :]; \Theta)$
        compute $L = loss(\hat{\mathcal{Y}}[:, : r, :], \mathcal{Y}[:, : r, id(V_i)])$
        compute the stochastic gradient of Θ according to L.
        update model parameters Θ according to their gradients and the learning rate γ.
    end for
    $iter = iter + 1$.
until convergence

Table 1: Dataset statistics.

Datasets	# Samples	# Nodes	Sample Rate	Input Length	Output Length
traffic	17,544	862	1 hour	168	1
solar-energy	52,560	137	10 minutes	168	1
electricity	26,304	321	1 hour	168	1
exchange-rate	7,588	8	1 day	168	1
metr-la	34272	207	5 minutes	12	12
pems-bay	52116	325	5 minutes	12	12

Table 2: Baseline comparison under single-step forecasting for multivariate time series methods.

Dataset		Solar-Energy				Traffic				Electricity				Exchange-Rate			
		Horizon				Horizon				Horizon				Horizon			
Methods	Metrics	3	6	12	24	3	6	12	24	3	6	12	24	3	6	12	24
AR	RSE	0.2435	0.3790	0.5911	0.8699	0.5991	0.6218	0.6252	0.63	0.0995	0.1035	0.1050	0.1054	0.0228	0.0279	0.0353	0.0445
	CORR	0.9710	0.9263	0.8107	0.5314	0.7752	0.7568	0.7544	0.7519	0.8845	0.8632	0.8591	0.8595	0.9734	0.9656	0.9526	0.9357
VARMLP	RSE	0.1922	0.2679	0.4244	0.6841	0.5582	0.6579	0.6023	0.6146	0.1393	0.1620	0.1557	0.1274	0.0265	0.0394	0.0407	0.0578
	CORR	0.9829	0.9655	0.9058	0.7149	0.8245	0.7695	0.7929	0.7891	0.8708	0.8389	0.8192	0.8679	0.8609	0.8725	0.8280	0.7675
GP	RSE	0.2259	0.3286	0.5200	0.7973	0.6082	0.6772	0.6406	0.5995	0.1500	0.1907	0.1621	0.1273	0.0239	0.0272	0.0394	0.0580
	CORR	0.9751	0.9448	0.8518	0.5971	0.7831	0.7406	0.7671	0.7909	0.8670	0.8334	0.8394	0.8818	0.8713	0.8193	0.8484	0.8278
RNN-GRU	RSE	0.1932	0.2628	0.4163	0.4852	0.5358	0.5522	0.5562	0.5633	0.1102	0.1144	0.1183	0.1295	0.0192	0.0264	0.0408	0.0626
	CORR	0.9823	0.9675	0.9150	0.8823	0.8511	0.8405	0.8345	0.8300	0.8597	0.8623	0.8472	0.8651	0.9786	**0.9712**	0.9531	0.9223
LSTNet-skip	RSE	0.1843	0.2559	0.3254	0.4643	0.4777	0.4893	0.4950	0.4973	0.0864	0.0931	0.1007	0.1007	0.0226	0.0280	0.0356	0.0449
	CORR	0.9843	0.9690	0.9467	0.8870	0.8721	0.8690	0.8614	0.8588	0.9283	0.9135	0.9077	0.9119	0.9735	0.9658	0.9511	0.9354
TPA-LSTM	RSE	0.1803	**0.2347**	0.3234	0.4389	0.4487	0.4658	0.4641	0.4765	0.0823	0.0916	0.0964	0.1006	**0.0174**	**0.0241**	**0.0341**	**0.0444**
	CORR	0.9850	**0.9742**	0.9487	0.9081	0.8812	0.8717	0.8717	0.8629	0.9439	0.9337	0.9250	0.9133	**0.9790**	0.9709	0.9564	0.9381
MTGNN	RSE	**0.1778**	0.2348	**0.3109**	**0.4270**	**0.4162**	0.4754	**0.4461**	**0.4535**	**0.0745**	0.0878	**0.0916**	**0.0953**	0.0194	0.0259	0.0349	0.0456
	CORR	**0.9852**	0.9726	**0.9509**	0.9031	**0.8963**	0.8667	**0.8794**	**0.8810**	**0.9474**	0.9316	**0.9278**	**0.9234**	0.9786	0.9708	0.9551	0.9372
MTGNN+sampling	RSE	0.1875	0.2521	0.3347	0.4386	0.4170	**0.4435**	0.4469	0.4537	0.0762	**0.0862**	0.0938	0.0976	0.0212	0.0271	0.0350	0.0454
	CORR	0.9834	0.9687	0.9440	0.8990	0.8960	**0.8815**	0.8793	0.8758	0.9467	**0.9354**	0.9261	0.9219	0.9788	0.9704	**0.9574**	**0.9382**

	Horizon 3			Horizon 6			Horizon 12		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
METR-LA									
DCRNN	2.77	5.38	7.30%	3.15	6.45	8.80%	3.60	7.60	10.50%
STGCN	2.88	5.74	7.62%	3.47	7.24	9.57%	4.59	9.40	12.70%
Graph WaveNet	2.69	5.15	6.90%	3.07	6.22	8.37%	3.53	7.37	10.01%
ST-MetaNet	2.69	5.17	6.91%	3.10	6.28	8.57%	3.59	7.52	10.63%
MRA-BGCN	**2.67**	**5.12**	6.80%	3.06	6.17	8.30%	3.49	7.30	10.00%
GMAN	2.77	5.48	7.25%	3.07	6.34	8.35%	**3.40**	**7.21**	**9.72%**
MTGNN	2.69	5.18	6.86%	**3.05**	**6.17**	**8.19%**	3.49	7.23	9.87%
MTGNN+sampling	2.76	5.34	**5.18%**	3.11	6.32	8.47%	3.54	7.38	10.05%
PEMS-BAY									
DCRNN	1.38	2.95	2.90%	1.74	3.97	3.90%	2.07	4.74	4.90%
STGCN	1.36	2.96	2.90%	1.81	4.27	4.17%	2.49	5.69	5.79%
Graph WaveNet	1.30	2.74	**2.73%**	1.63	3.70	3.67%	1.95	4.52	4.63%
ST-MetaNet	1.36	2.90	2.82%	1.76	4.02	4.00%	2.20	5.06	5.45%
MRA-BGCN	**1.29**	**2.72**	2.90%	**1.61**	**3.67**	3.80%	1.91	4.46	4.60%
GMAN	1.34	2.82	2.81%	1.62	3.72	**3.63%**	**1.86**	**4.32**	**4.31%**
MTGNN	1.32	2.79	2.77%	1.65	3.74	3.69%	1.94	4.49	4.53%
MTGNN+sampling	1.34	2.83	2.83%	1.67	3.79	3.78%	1.95	4.49	4.62%

HetETA: Heterogeneous Information Network Embedding for Estimating Time of Arrival

Huiting Hong*
Yucheng Lin*
honghuiting@didiglobal.com
linyucheng@didiglobal.com
AI Labs, Didi Chuxing

Xiaoqing Yang
Zang Li
yangxiaoqing@didiglobal.com
lizang@didiglobal.com
AI Labs, Didi Chuxing

Kun Fu
fukunkunfu@didiglobal.com
AI Labs, Didi Chuxing

Zheng Wang
wangzhengzwang@didiglobal.com
AI Labs, Didi Chuxing

Xiaohu Qie
tiger.qie@didiglobal.com
Technology Ecosystem & Development, Didi Chuxing

Jieping Ye
yejieping@didiglobal.com
AI Labs, Didi Chuxing

Background

- Estimated Time of Arrival (ETA)： 根据给定的出发时间，精确预估从起点到终点所需时。有助于节省用户出行时间，优化车辆调度和路径规划等。

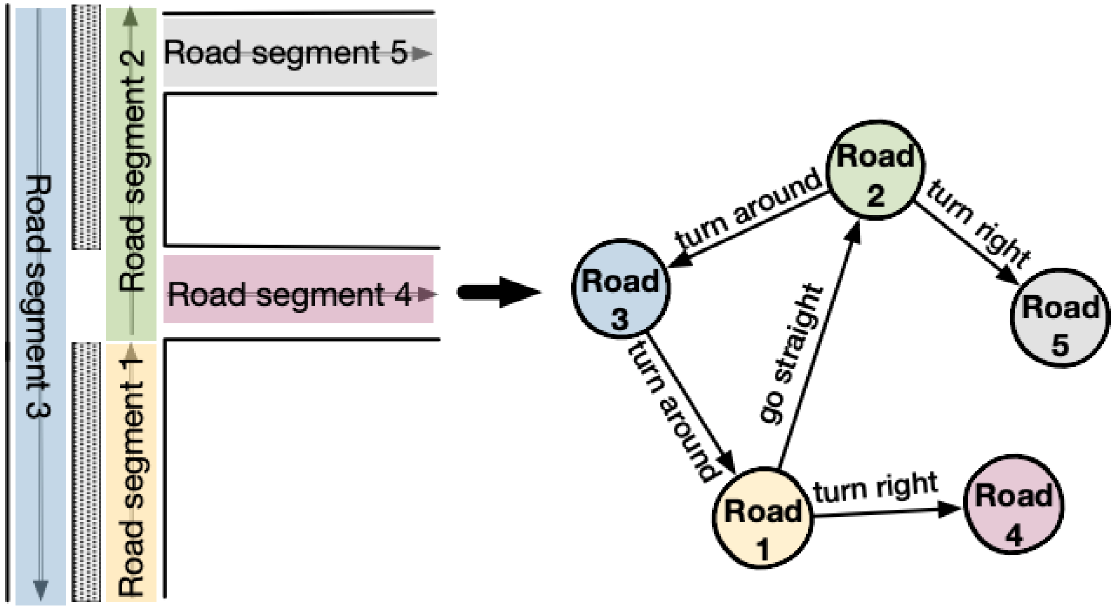
Motivation

1、当前大多数工作致力于建立丰富的特征系统来提高 ETA 任务的准确性，然而这些特征系统很少考虑到空间信息的构建与挖掘。

2、道路网络是一个大规模的稀疏网络，这种大规模稀疏网络难以直接用需要充足邻居信息的GNN进行学习。

3、ETA任务与交通速度的时序信息有关。

4、除了地图数据中的道路网络，车辆轨迹信息也是一种描述道路之间链接关系的空间信息。



Road map Multi-relational road network

Contribution

1、提出了HetETA框架使用卷积和图神经网络从空间（车流、道路）和时间（最近、每日、每周）提取出异质信息。

2、设计了一个基于注意的具有快速局部谱滤波的图网络Het-ChebNet，在与边数成正比的空间需求下嵌入稀疏异构信息网络。

3、设计了大量实验证明了网络的有效性。

Problem Statement

给定出发的查询q，t_q 表示出发时间，O_q表示出发地点，d_q代表目的地点，Pq代表路径

$$q = (o_q, d_q, t_q, P_q)$$

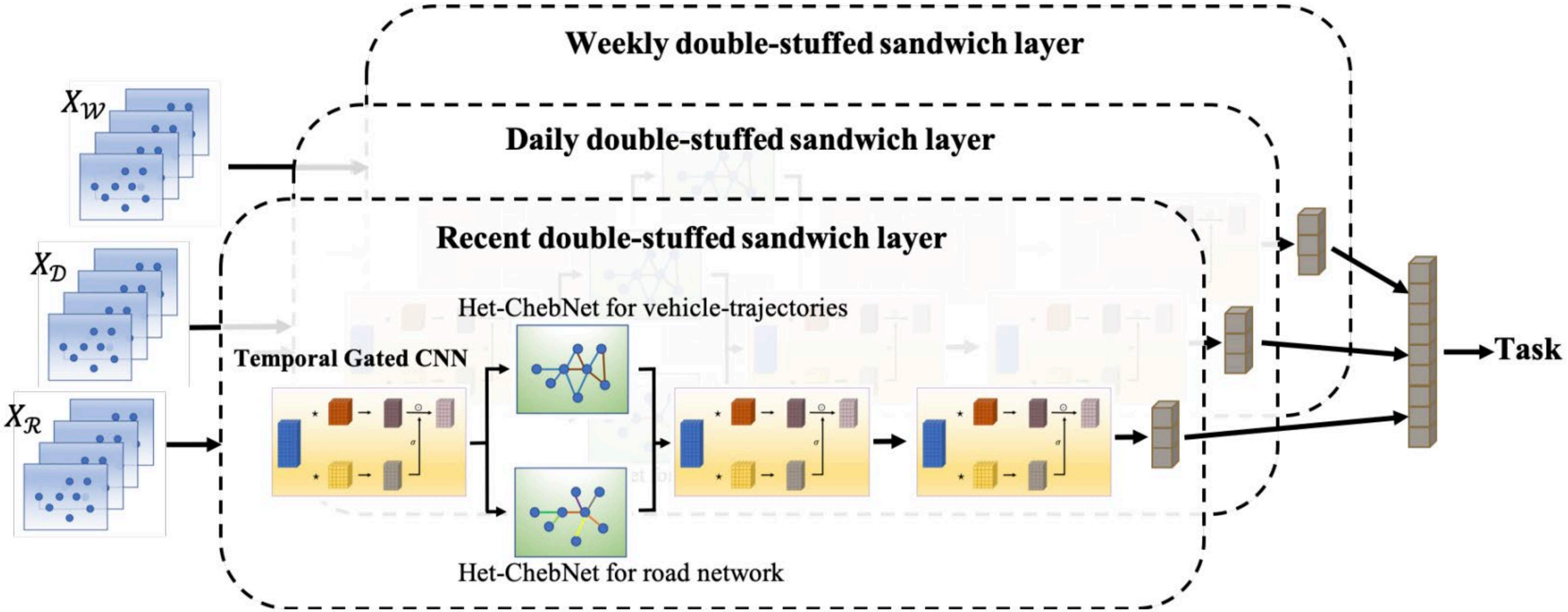
交通网络集合为：

$$\{G^{(t_q-\tau+1)}, G^{(t_q-\tau+2)}, \ldots, G^{(t_q)}\}$$

$$G^{(t)} = (V, E, R, X^{(t)})$$

$$e_{ijk} = (v_i, v_j, r_k) \in E$$

Methods



Weekly double-stuffed sandwich layer
Daily double-stuffed sandwich layer
Recent double-stuffed sandwich layer
$X_\mathcal{W}$
$X_\mathcal{D}$
$X_\mathcal{R}$
Het-ChebNet for vehicle-trajectories
Temporal Gated CNN
Het-ChebNet for road network
Task

Methods

- 网络由三个相同结构组件构成，每个组件分别用于处理不同时序路况

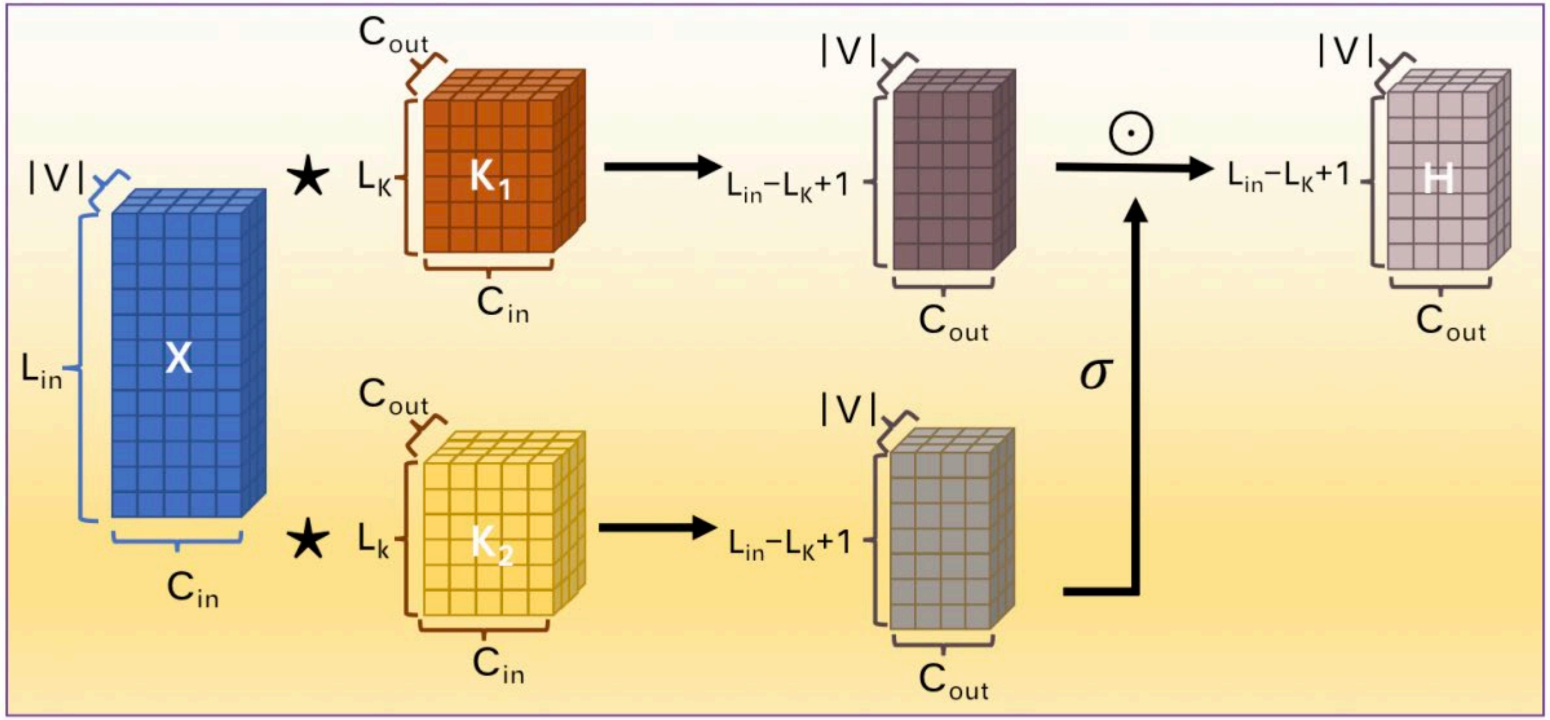
$$\mathrm{X}_{\mathcal{R}} = \left[X^{(t_q-L_{\mathcal{R}}+1)},\ X^{(t_q-L_{\mathcal{R}}+2)}, \ldots, X^{(t_q)}\right] \in \mathbb{R}^{L_{\mathcal{R}}\times|V|\times n}$$

$$\mathrm{X}_{\mathcal{D}} = \left[X^{(t_q+1-L_{\mathcal{D}}*T_{\mathcal{D}})},\ X^{(t_q+1-(L_{\mathcal{D}}-1)*T_{\mathcal{D}})}, \ldots, X^{(t_q+1-T_{\mathcal{D}})}\right] \in \mathbb{R}^{L_{\mathcal{D}}\times|V|\times n}$$

$$\mathrm{X}_{\mathcal{W}} = \left[X^{(t_q+1-L_{\mathcal{W}}*T_{\mathcal{D}}*7)},\ X^{(t_q+1-(L_{\mathcal{W}}-1)*T_{\mathcal{D}}*7)},\ \ldots,\ X^{(t_q+1-T_{\mathcal{D}}*7)}\right] \in \mathbb{R}^{L_{\mathcal{W}}\times|V|\times n}$$

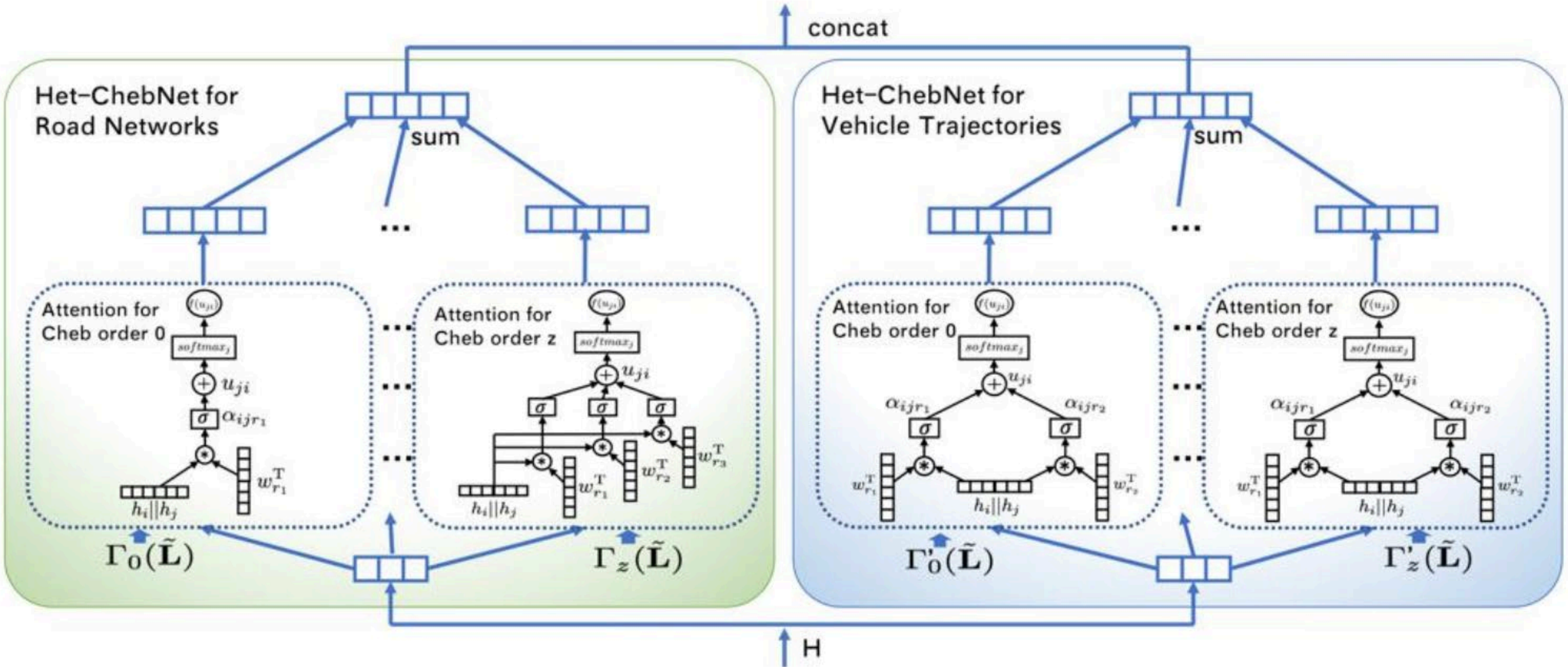
$$\hat{\mathrm{Y}} = \frac{\mathbf{S}}{\sigma\left(\left[\mathrm{H}_{\mathcal{R}4}||\mathrm{H}_{\mathcal{D}4}||\mathrm{H}_{\mathcal{W}4}\right]\mathbf{W}+b\right)*120}$$

Methods



$$\mathbf{H} = (\mathbf{K}_1 \star \mathbf{x}) \odot \sigma(\mathbf{K}_2 \star \mathbf{x}) \in \mathbb{R}^{(L_{\text{in}} - L_K + 1) \times |V| \times C_{\text{out}}},$$

Methods



concat
Het-ChebNet for
Road Networks
sum
Het-ChebNet for
Vehicle Trajectories
sum
Attention for
Cheb order 0
Attention for
Cheb order z
Attention for
Cheb order 0
Attention for
Cheb order z
u_{ji}
α_{ijr_1}
α_{ijr_2}
$h_i||h_j$
$w_{r_1}^{T}$
$w_{r_2}^{T}$
$w_{r_3}^{T}$
$\Gamma_0(\tilde{\mathbf{L}})$
$\Gamma_z(\tilde{\mathbf{L}})$
$\Gamma'_0(\tilde{\mathbf{L}})$
$\Gamma'_z(\tilde{\mathbf{L}})$
H

Methods

车流网络所构建的异质图一定程度上缓解了道路网络的稀疏问题，为了克服基于 GCN的图卷积模型无法在稀疏的道路网络上收集充足的邻居信息的问题，该研究采用了基于谱图理论的 ChebNet 网络，通过切比雪夫多项式构成局域滤波器：

$$g_\theta \star_G \mathbf{x} = g_\theta(\mathbf{L})\mathbf{x} = \sum_{z=0}^{Z} \theta_z \Gamma_z(\tilde{\mathbf{L}})\mathbf{x},$$

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}} \in \mathbb{R}^{|V|\times|V|}$$

$$\tilde{\mathbf{L}} = 2\mathbf{L}/\lambda_{\max} - \mathbf{I},$$

$$\Gamma_z(\tilde{\mathbf{L}}) = 2\tilde{\mathbf{L}}\Gamma_{z-1}(\tilde{\mathbf{L}}) - \Gamma_{z-2}(\tilde{\mathbf{L}}) \text{ with } \Gamma_0(\tilde{\mathbf{L}}) = \mathbf{I} \text{ and } \Gamma_1(\tilde{\mathbf{L}}) = \tilde{\mathbf{L}}.$$

Methods

然而，传统的 ChebNet 网络无法处理异质图中所包含的多关系信息，因此该研究基于 ChebNet 提出了一个能够捕捉多关系链接信息的 Het-ChebNet：

$$g_\theta \star_G \mathbf{x} = g_\theta(\mathbf{L})\mathbf{x} = \sum_{z=0}^{Z} \theta_z \left(\Gamma_z(\tilde{\mathbf{L}}) \odot f(\mathbf{U}_z) \right) \mathbf{x},$$

$$\alpha_{ijk} = \sigma \left({w_{r_k}}^{\mathrm{T}} \left[h_i || h_j \right] \right),$$

$$u_{ji} = \sum_{k,(v_i,v_j,r_k) \in E} \alpha_{ijk}.$$

$$f(u_{ji}) = \exp(u_{ji}) / \sum_{k,(v_k,v_j) \in E} \exp(u_{jk}).$$

EXPERIMENTS

Dataset	SY_6_Trip			SY_6_Pickup		
Metric	MAPE	MAE	RMSE	MAPE	MAE	RMSE
GRU	13.84%	129.99	216.52	24.89%	52.37	91.86
DCRNN	13.21%	124.01	208.10	24.09%	49.62	85.69
STGCN	12.88%	119.96	200.76	23.33%	47.55	**82.24**
GWN*	12.89%	121.60	205.39	23.64%	48.96	85.54
ASTGCN*	12.57%	117.53	119.17	23.39%	48.50	86.13
HetETA	**12.32%**	**116.44**	**197.26**	**22.96%**	**47.16**	82.77
Dataset	SY_7_Trip			SY_7_Pickup		
Metric	MAPE	MAE	RMSE	MAPE	MAE	RMSE
GRU	13.99%	123.27	193.06	23.77%	51.75	83.28
DCRNN	13.28%	116.80	183.48	23.04%	49.09	77.62
STGCN	12.94%	113.48	178.42	22.29%	46.59	74.01
GWN*	13.01%	114.08	179.25	22.60%	47.66	75.98
ASTGCN*	12.66%	111.26	175.95	22.28%	47.22	76.27
HetETA	**12.39%**	**109.17**	**173.03**	**21.89%**	**46.23**	**73.78**