Disentanglement-Guided Spatial-Temporal Graph Neural Network for Metro Flow Forecasting (Student Abstract)

Jinyu Hong[1], Ping Kuang[1], Qiang Gao[2,3*], Fan Zhou[1,3]
[1]University of Electronic Science and Technology of China
[2]Southwestern University of Finance and Economics
[3]Kash Institute of Electronics and Information Industry
qianggao@swufe.edu.cn,{jinyuhong@std.,kuangping@,fan.zhou@}uestc.edu.cn

Abstract

In recent intelligent transportation applications, metro flow forecasting has received much attention from researchers. Most prior arts endeavor to explore spatial or temporal dependencies while ignoring the key characteristic patterns underlying historical flows, e.g., trend and periodicity. Although the multiple granularity distillations or spatial dependency correlation can promote the flow estimation. However, the potential noise and spatial dynamics are under-explored. To this end, we propose a novel Disentanglement-Guided Spatial-Temporal Graph Neural Network or DGST to address the above concerns. It contains a Disentanglement Pre-training procedure for characteristic pattern disentanglement learning, a Characteristic Pattern Prediction for different future characteristic explorations, and a Spatial-Temporal Correlation for spatial-temporal dynamic learning. Experiments on a real-world dataset demonstrate the superiority of our DGST.

Introduction

Metro flow forecasting serves as a pivotal task within the realm of Intelligent Transportation Systems as it plays a paramount role in effectively managing passenger flow and optimizing metro scheduling. Recent efforts on flow learning have devised diverse deep neural networks such as RNN variants to model temporal dynamics as well as developing graph networks to incorporate explicit spatial dependencies.

Despite the effectiveness, existing efforts still have three major limitations for metro flow prediction. First, historical metro flows themselves have a variety of temporal characteristics or signals, such as trend and periodicity, which play different roles in future flow evolutions. However, existing solutions are usually designed locally and most of them use the flow sequence generated from the past one hour as input data, which results in a narrow perspective to capture these temporal characteristics and may lead to inaccurate forecasts. Second, some works attempt to learn periodicity features from different temporal granularities individually to uncover the inherent flow dynamics. For instance, ASTGCN (Guo et al. 2019) categorizes historical flow into hourly and weekly granularities to enhance periodicity pattern learning. However, such approaches could result in noise from hand-crafted flow segments, thereby leading to inaccurate estimations. Third, existing practices mainly apply a static adjacency matrix to learn spatial relations. However, the spatial dependencies between stations are highly dynamic. Hence, they lack the ability to model time-varying spatial dependencies.

To address the limitations above, we propose a novel **D**isentanglement-**G**uided **S**patial-**T**emporal graph network for metro flow forecasting or DGST. DGST has a disentanglement pre-training procedure, which uses a devised variational disentanglement to decompose trend and periodicity signals into different latent spaces, which enables more interpretable and robust representations from historical metro flows. Next, we design a Characteristic Pattern Prediction module to generate the future characteristics of trend and periodicity, respectively. To uncover the dynamics underlying spatial dependencies, we build a spatial-temporal correlation module with an attention mechanism to adaptively determine the station relations. Finally, we fuse the outputs from the above modules to make future flow forecasting.

Methodology

Problem Formulation. Let $\boldsymbol{X}_t \in \mathbb{R}^{N\times D}$ denote the metro flow of N stations at time t, where D represents the number of observation types. Given historical metro flow $\boldsymbol{X}_\Omega \in \mathbb{R}^{T_h\times N\times D}$, our goal is to learn a function f_θ from $\boldsymbol{X}_\Omega$ to forecast T steps' flow in the future, denoted as:

$$[\boldsymbol{X}_{t+1}, \cdots, \boldsymbol{X}_{t+T}] = f_\theta[\boldsymbol{X}_{t-T_h+1}, \cdots, \boldsymbol{X}_t]. \quad (1)$$

Disentanglement Pre-training. Inspired by previous study (Wang et al. 2022), we devise a disentangled encoder by following the rule of variational Bayes and information theory. Given historical metro flow $\boldsymbol{X}_\Omega$, we employ two fully connected layers as the encoder skeleton to respectively model periodicity posterior $q(z^p|\boldsymbol{X}_\Omega)$ and trend posterior $q(z^t|\boldsymbol{X}_\Omega)$, where the prior assumptions (e.g., $p(z^t)$) are derived from the standard Gaussian distribution. Then, we introduce an approximate estimator with Jensen-Shannon divergence to make posterior disengagement in a self-supervised manner, where the objective is to minimize:

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{rec} \quad (2)\\ &- \mathcal{KL}\left[q\left(z^t \mid \boldsymbol{X}_\Omega\right) \| p(z^t)\right] - \mathcal{KL}\left[q\left(z^p \mid \boldsymbol{X}_\Omega\right) \| p(z^p)\right]\\ &+ \mathbb{E}_{q(z^t|z^p)}[-\mathrm{s}(-T_\theta(z^t, z^p))] - \mathbb{E}_{q(z^t)q(z^p)}[\mathrm{s}(T_\theta(z^t, z^p)].\end{aligned}$$

*Corresponding Author (qianggao@swufe.edu.cn).

Herein, $\mathcal{L}_{rec}$ refers to reconstruction loss, $\mathcal{KL}$ denotes KL divergence, $s(\cdot)$ is the softplus function, and T_θ is a discriminator function modeled by a fully connected network.

Characteristic Pattern Prediction. We aim to use the pre-trained encoder to generate z^t and z^p from $\boldsymbol{X}_\Omega$, respectively. For trend characteristic, we feed z^t to stacked MLPs to predict future trend $\boldsymbol{Y}_t \in \mathbb{R}^{T\times N\times D}$.In cases where the metro flow exhibits strong seasonality, frequency-domain models tend to be more sample-efficient in practice than time-domain attention models. Hence, we apply the Fourier Transform to tackle z^p by performing an attention network in the frequency domain and then convert the results back to the time domain using the inverse Fourier transform. Finally, we can generate the future periodicity $\boldsymbol{Y}_p \in \mathbb{R}^{T\times N\times D}$.

Spatial-Temporal Correlation. Instead of using a static adjacency matrix, we employ the input of each time step as the key and query in self-attention to construct a fully connected graph, which represents the dynamic interactions between metro stations. We assume that regional heterogeneity can be estimated by the difference between their long-term temporal patterns. Consequently, to maintain regional heterogeneity, we introduce two adjacency matrices, denoted as A^t and A^p, whose elements are constructed based on the similarity of distributions of long-term trend and periodicity. We consider the above matrices as additional signals to facilitate the learning of spatial-temporal correlations, where the attention score of $\boldsymbol{X}_t$ can be computed as:

$$\boldsymbol{\alpha}_t = \mathcal{S}(\frac{(W_q\boldsymbol{X}_t)(W_k\boldsymbol{X}_t)^T}{\sqrt{d}} + W_t \odot A^t + W_p \odot A^p), \quad (3)$$

where $\mathcal{S}$ denotes the Softmax function, W_* are learnable parameters, $\odot$ is the element-wise Hadamard product. We apply 1^{st} order Chebyshev polynomial as graph convolution operation and replace the MLPs in GRU with spatial graph convolution to capture spatial and temporal correlations.

Fusion & Forecasting. We design a fusion layer to effectively unite the learned components and make forecasting:

$$\boldsymbol{X}_{t+1}, \cdots, \boldsymbol{X}_{t+T} = \boldsymbol{W}_{ST} \odot \boldsymbol{Y} + \boldsymbol{W}_{LT} \odot (\boldsymbol{Y}_t + \boldsymbol{Y}_p), \quad (4)$$

where $\boldsymbol{Y}$ is the output of the Spatial-Temporal Correlation module, and $\boldsymbol{W}_*$ are learnable parameters.

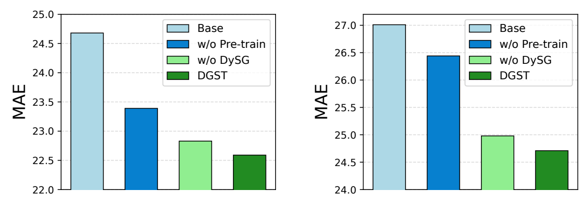
Experimental Results

We conduct experiments on Hangzhou Metro dataset and compare with the following popular baselines: STGCN (Yu, Yin, and Zhu 2018), ASTGCN (Guo et al. 2019), AGCRN (Bai et al. 2020), STGODE (Fang et al. 2021), MegaCRN (Jiang et al. 2023). We adopt Mean Absolute Error(MAE), Root Mean Square Error(RMSE) and Mean Absolute Percentage Error(MAPE) as our evaluation metrics.

Table 1 summarizes the experimental results. In particular, DGST outperforms all baselines in both short-term and long-term forecasts. These results demonstrate the benefits of DGST in capturing spatial-temporal dependencies for metro flow forecasting. Besides, we compare DGST with the base model, the model without pre-train (w/o Pre-train) and the model without dynamic spatial graph (w/o DySG). Fig 1 shows the results of these variants on the inflow aspect.

Model	**Metric**	**Inflow (min)**			**Outflow (min)**		
		15	30	60	15	30	60
STGCN	MAE	24.20	25.73	29.54	25.42	28.18	35.34
	RMSE	40.61	43.60	51.83	43.44	47.84	60.16
	MAPE	0.213	0.221	0.249	0.181	0.201	0.247
ASTGCN	MAE	26.16	26.30	26.91	26.77	26.79	27.45
	RMSE	46.70	46.92	48.88	43.97	44.12	46.02
	MAPE	0.251	0.261	0.294	0.196	0.201	0.232
AGCRN	MAE	23.24	24.09	26.05	21.90	22.90	**24.52**
	RMSE	39.26	41.20	46.33	39.32	40.54	43.08
	MAPE	0.204	0.210	0.230	0.149	0.159	0.172
STGODE	MAE	27.09	29.15	35.38	27.46	29.94	36.26
	RMSE	46.25	50.09	64.09	49.06	53.46	66.80
	MAPE	0.247	0.271	0.377	0.197	0.220	0.289
MegaCRN	MAE	25.83	26.88	29.85	27.33	29.01	32.44
	RMSE	44.81	47.49	58.55	51.26	55.11	63.32
	MAPE	0.214	0.214	0.253	0.201	0.213	0.247
DGST	MAE	**22.59**	**23.34**	**24.71**	**21.78**	**22.51**	24.62
	RMSE	**38.56**	**40.30**	**44.40**	**37.11**	**38.23**	**41.84**
	MAPE	**0.199**	**0.202**	**0.214**	**0.147**	**0.152**	**0.168**

Table 1: Performance comparisons.



(a) MAE of 15min ahead. (b) MAE of 60min ahead.

Figure 1: Ablation Study on Hangzhou Metro.

Acknowledgments

This work was supported by the Natural Science Foundation of Sichuan Province under Grant No.2023NSFSC1411 and the National Natural Science Foundation of China under Grant No.62102326, No.62176043, and No.62072077.

References

Bai, L.; Yao, L.; Li, C.; Wang, X.; and Wang, C. 2020. Adaptive graph convolutional recurrent network for traffic forecasting. *NIPS*, 33: 17804–17815.

Fang, Z.; Long, Q.; Song, G.; and Xie, K. 2021. Spatial-temporal graph ode networks for traffic flow forecasting. In *KDD*, 364–373.

Guo, S.; Lin, Y.; Feng, N.; Song, C.; and Wan, H. 2019. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *AAAI*, volume 33, 922–929.

Jiang, R.; Wang, Z.; Yong, J.; Jeph, P.; Chen, Q.; Kobayashi, Y.; Song, X.; Fukushima, S.; and Suzumura, T. 2023. Spatio-temporal meta-graph learning for traffic forecasting. In *AAAI*, 7, 8078–8086.

Wang, Z.; Xu, X.; Zhang, W.; Trajcevski, G.; Zhong, T.; and Zhou, F. 2022. Learning latent seasonal-trend representations for time series forecasting. *NIPS*, 35: 38775–38787.

Yu, B.; Yin, H.; and Zhu, Z. 2018. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In *IJCAI*, 3634–3640.