Proceedings of CONF-MLA 2024 Workshop: Mastering the Art of GANs: Unleashing Creativity with Generative Adversarial Networks
DOI: 10.54254/2755-2721/111/2024CH0106

A Cross-City Traffic Flow Prediction Framework Incorporating Holidays and Weather Factors

Shuhang Cui

School of Physics and Astronomy Sciences, Sun Yat-Sen University, Zhuhai, China

cuishh@mail2.sysu.edu.com

Abstract. Although graph neural networks are widely used in urban traffic flow prediction research, they often face challenges related to data scarcity. Cross-city transfer learning presents a promising and cost-effective solution to this problem. This study introduces TransGTR-MCA (Meteorological and Calendrical Aware), an advanced model for cross-city traffic flow prediction based on the TransGTR framework. TransGTR-MCA utilizes a Graph Structure Learning framework that creates adaptive graph structures dynamically, without depending on predefined adjacency matrices. This method enhances the ability to capture complex spatial dependencies, especially in cross-city contexts. A key innovation of the model is its incorporation of external factors like weather and holidays, which allows for a thorough consideration of the unique climatic and cultural traits of different cities. This integration aims to improve the transfer learning process across various urban settings. To assess the model's effectiveness, experiments were carried out using three benchmark datasets for traffic flow prediction: METR-LA, PEMSD7M, and SZ-TAXI. The results demonstrate that the TransGTR-MCA model outperforms others in short-term predictions. Specifically, when transferring knowledge from the U.S. METR-LA dataset to the Chinese SZ-TAXI dataset, the model achieved a notable 2.73% reduction in error for 5-minute predictions, highlighting its potential for cross-cultural urban traffic forecasting.

Keywords: Transfer learning, Spatio-temporal Prediction, Graph convolutional network.

1. Introduction

Traffic flow prediction is a critical task in smart city management, often facing data scarcity challenges. As urban populations grow and transportation networks become increasingly complex, accurate predictions are essential for efficient traffic control, resource allocation, and urban planning. Most urban data not only change over time but also exhibit correlations and heterogeneity across different spatial locations. To capture this complex dual characteristic in urban traffic networks, Spatial-Temporal Graph Neural Networks (STGNNs) have emerged, combining temporal models and graph neural networks to simultaneously process spatiotemporal traffic data characteristics [1].

More specifically, there are two mainstream strategies: factorized or coupled neural architecture. The representative model of the former is the Spatial-Temporal Graph Convolutional Network (STGCN) [2], which processes temporal features and spatial information in parallel and utilizes convolutional structures to learn multi-scale features. Despite the advanced capabilities of STGNNs, data scarcity in traffic prediction remains a significant challenge, particularly in newly developed urban areas or cities with limited sensing infrastructure. Transfer learning is an important solution to this problem. Cross-city transfer learning based on meta-learning frameworks has been applied in the field of traffic prediction [3], improving prediction accuracy in data-scarce scenarios by transferring knowledge between cities. This approach allows cities with limited data to benefit from knowledge gained in data-rich environments, potentially accelerating the deployment of intelligent transportation systems in developing urban areas. These methods provide promising solutions for improving the performance of traffic prediction models in data-limited environments, paving the way for more adaptive and resilient urban transportation systems.

In traffic flow prediction research, it is crucial to consider the impact of weather factors and public holidays across different geographical and cultural contexts. Precipitation events significantly affect traffic flow. Studies show that in London, even light rainfall (less than 0.1 mm/h) can increase travel time by 0.1-2.1%[4]. In the United States, light rain conditions can reduce vehicle speed by 2-3.6% [5]. In China, rainy weather can lead to a 5.4% decrease in weighted road speed [6]. These studies reflect that the impact of adverse weather on traffic also exhibits regional differences. Integrating accurate meteorological data into traffic prediction models is essential. For example, a Weather Interaction-Aware Spatio-Temporal Attention Network (WST-ANet) model that considers weather factors has already achieved superior performance on the PeMS04 and PeMS08 datasets [7].

On the other hand, public holidays also significantly influence traffic patterns and exhibit certain temporal effects and traffic preference characteristics [8]. Research on Chinese cities confirms that during non-holiday periods, the net intra-city flow shows periodic dynamics with notable peaks on Mondays and Saturdays. During holiday periods, net flows experience significant spatiotemporal fluctuations, leading to cities being classified as inbound, outbound, or balanced types. Sequence forecasting based on convolutional neural networks and long short-term memory has already demonstrated superior performance in predicting peak flows in specific cases [10]. However, it is important to note that different cultural backgrounds and industrial structures may lead to significant regional differences in holiday effects.

TransGTR is the latest cross-city transfer learning framework [11], which couples mature deep learning models such as TSFormer [12], GTS [13], and Graph-WaveNet [14]. It attempts to address issues in traditional transfer learning such as insufficient model generalization and sparse data in target cities through techniques like knowledge distillation and spatio-temporal decoupling. Compared to baseline models, it has achieved significant optimization effects. The distinctive feature of this framework is that it does not use predefined graph structures, but instead builds the model using a structure generator starting from the node feature network, effectively capturing the inherent dynamic characteristics of urban traffic networks.

This paper proposes TransGTR-MCA (Meteorological and Calendrical Aware), an improved model based on TransGTR, further coupling external factors such as meteorological information and holiday information to enhance the model's transfer learning efficiency across climate zones and cultural areas. It will also address several related issues, further extrapolating the model to real datasets in mainland China, and explore the application value of the TransGTR-MCA model.

2. Method

2.1. Capturing short to long-term temporal features

The features of time series can be artificially divided into three parts: long-term dependencies beyond a week, short-term dependencies under 1 hours, and other potentially influential features (such as daily cycles). The TSFormer model uses a fixed patch size, then applies 2D convolution operations to embed each patch into a high-dimensional space, and finally randomly masks the global data. During each learning process, the model simultaneously learns both global and local features, achieving the simplest form of multi-scale learning in the time dimension. Other temporal features are indirectly represented.

DOI: 10.54254/2755-2721/111/2024CH0106

2.2. Handling external factors

In addition to traffic flow data, we matched and supplemented the sequence dataset with day-of-week encoding, holiday encoding, and weather data encoding. The holiday data was sourced from the Python package 'holidays', while the weather data was obtained from the Global Surface Summary of the Day (GSOD) records maintained by NOAA (National Oceanic and Atmospheric Administration). For simplicity, we focused on studying the potential impact of precipitation. For each time step, if a precipitation event occurred, it was marked as 1.

2.3. Pretraining and knowledge distillation

The pretraining process is implemented using Transformer-based code. For each spatial node or variable, the TSFormer model first learns in the temporal dimension, capturing both local and global features, but does not perform cross-node computations. The calculation of self-attention scores follows the standard method, after which Softmax is applied to obtain attention weights, followed by a linear layer and dropout layer.

Knowledge distillation is a powerful technique for transferring knowledge from a large, complex model (the teacher) to a smaller, more efficient model (the student). This framework directly uses the unsoftened teacher model outputs for knowledge distillation, conveying more certain knowledge, which avoids the potential over-smoothing caused by high temperatures in regression prediction tasks. Additionally, when calculating the difference between the teacher and student models, the program uses L2 distance (mean squared error) instead of Kullback-Leibler (KL) divergence, further preserving precise numerical information. Finally, in the optimization phase, a hyperparameter λ_d controls the weight of the distillation model in the framework.

$$L_{distil} = L_{source} + L_{target} \tag{1}$$

$$\min L_{MAE} + \lambda_d L_{distil} \tag{2}$$

By default, it equally emphasizes learning features from the teacher model and the data. The process allows the student to achieve performance comparable to the teacher while significantly reducing model size and inference time.

2.4. Structure generation and transfer learning

TransGTR designs a framework that couples the GTS model to generate adaptive graph structures. The GTS model can dynamically learn and adjust relationships between nodes without relying on predefined graph structures. This approach excels in capturing complex spatial dependencies, especially when dealing with cross-city data. At the core of the GTS model is an attention-based structure generator that calculates relevance scores between node pairs and transforms these scores into adjacency matrix weights using Gumbel-Softmax sampling.

In terms of transfer learning, we utilize a meta-learning framework aimed at improving model performance in data-scarce situations. This framework consists of a meta-training phase and a meta-testing phase. During meta-training, we use rich data from multiple source cities to train a meta-model. This meta-model learns how to quickly adapt to new tasks or domains. In the meta-testing phase, we apply the trained meta-model to the target city, requiring only a small amount of local data for rapid adaptation.

$$L_{source} = L_{pred} + \lambda_1 L_{decreg} + \lambda_2 L_{CORAL} + \lambda_3 L_{recons} \tag{3}$$

The loss function comprises several components: L_{pred} is the prediction loss that measures the accuracy of the model's predictions; L_{decreg} is the degree regularization loss, which controls the sparsity of the generated graph; L_{CORAL} is the correlation alignment loss, aimed at reducing the discrepancy between the feature distributions of the source and target domains; and L_{recons} is the reconstruction loss, ensuring that the learned features can reconstruct the original input. The terms λ_1, λ_2 and λ_3 are weight coefficients used to balance the importance of different loss components in the overall objective function.

DOI: 10.54254/2755-2721/111/2024CH0106

On the other hand, the loss function for target domain fine-tuning is simplified to:

$$L_{target} = L_{pred} + \lambda_r L_{decreg} \tag{4}$$

where L_{pred} and L_{decreg} have the same definitions as in the source domain training, and λ_r is the weight coefficient for the degree regularization loss in the target domain.

Through this transfer learning method, the model can better adapt to the specific patterns of the target city while maintaining the reasonableness of the graph structure. This approach is particularly suitable for situations where data from the target city is limited, significantly improving the model's performance in new environments.

This framework combines dynamic graph structure generation and effective knowledge transfer, providing a powerful and flexible solution for cross-city traffic prediction tasks. And the model structure is illustrated in figure 1.

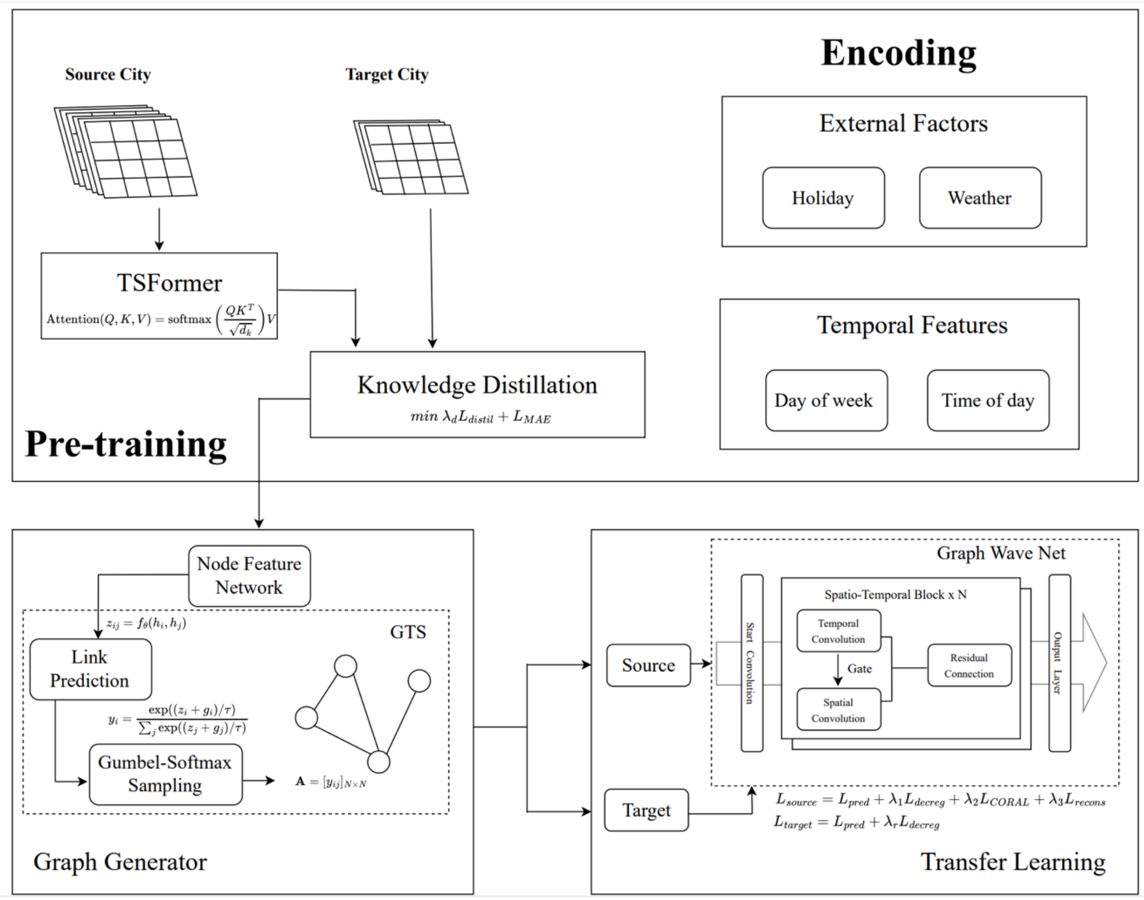

Figure 1. Schematic diagram of the model structure.

3. Result and discussion

This study selected widely used open-source datasets in the field of traffic flow prediction, namely METR-LA [15], PEMSD7M [16], and SZ-TAXI [17], as test subjects. These datasets represent different urban environments and traffic patterns, providing a diverse testing basis for model performance evaluation.

The METR-LA dataset, derived from the Los Angeles highway system, contains traffic speed data collected by 207 sensors over a 4-month period [1]. This dataset is characterized by its wide coverage and long-time span, capable of reflecting the complex traffic network dynamics of a large city. The

DOI: 10.54254/2755-2721/111/2024CH0106

PEMSD7M dataset, from the California Department of Transportation's Performance Measurement System (PeMS), includes traffic flow data from 228 detection stations over a 7-month period [2]. The advantage of this dataset lies in its high-quality data and rich time series information, suitable for long-term traffic trend analysis. The SZ-TAXI dataset collected GPS trajectory data from taxis in Shenzhen over one month, covering 156 test points across the entire city's traffic flow [3].

This paper uniformly selects the most comprehensive METR-LA as the source city dataset and examines its transfer learning efficiency on PEMSD7M and SZ-TAXI respectively. To maximally test the model's adaptability to data-scarce conditions, this paper consolidates the target city data length to 3 days. The comparison between the model considering weather factors and holiday factors and the original TransGTR model is shown in the table 1.

Due to differences in dataset scale, the SZ-TAXI dataset prediction results show smaller MAE and RMSE, but would have higher MAPE if calculated, revealing its data instability. Performance analysis shows that TransGTR-MCA performs best in most cases. For 5-minute prediction horizons, TransGTR-MCA outperforms TransGTR in both MAE and RMSE metrics. For 60-minute predictions, TransGTR-MCA is slightly better in MAE but slightly worse in RMSE compared to TransGTR.

Table 1. Evaluation of Model Prediction Performance base on 3-days data (Source dataset is METR-LA).

Horizon	**Metrics**	**PEMSD7M**		**SZ-TAXI**	
		TransGTR	TransGTR-MCA	TransGTR	TransGTR-MCA
5min	MAE	1.8197	1.3999	1.0067	0.9640
	RMSE	3.4313	2.3501	1.5236	1.4820
60min	MAE	6.1879	3.9077	2.8303	2.8110
	RMSE	10.2138	7.4434	4.0367	4.5564

In terms of model stability, TransGTR-MCA performs consistently in short-term predictions but shows a slight decline in RMSE for long-term predictions. In contrast, TransGTR demonstrates more consistent performance in both short and long-term predictions, especially in RMSE metrics, showing better stability.

Overall, the prediction difficulty increases significantly as the time range extends, with the error for 60-minute predictions reaching several times that of 5-minute predictions. TransGTR-MCA shows better enhancement for short-term forecasts but somewhat discounted results for long-term forecasts. This reflects the model's persistent limitations in capturing multi-feature long-term changes. Potential improvement suggestions include further strengthening the time series processing model and introducing attention network modules that better adapt to external factors.

4. Conclusion

This study successfully improved the TransGTR model's performance in cross-city traffic flow prediction by integrating weather and holiday factors. Experimental results demonstrate that the enhanced model significantly outperforms the original TransGTR in short-term predictions, particularly within the 5-minute forecast range. However, for long-term predictions, especially in the 60-minute RMSE metric, the model's optimization effect is slightly insufficient, reflecting the challenge of capturing complex long-term changes. This finding highlights the necessity for further optimization of time series processing and the introduction of attention mechanisms more adaptable to external factors. Future research should focus on enhancing the model's ability to capture long-term multidimensional feature changes to improve prediction accuracy and stability.

References

[1] Jin G, Liang Y, Fang Y, Shao Z, Huang J, Zhang J and Zheng Y (2023) Spatio-temporal graph neural networks for predictive learning in urban computing: A survey IEEE Trans. Knowl. Data Eng. https://doi.org/10.1109/TKDE.2023.3333824

[2] Yu B, Yin H and Zhu Z (2017) Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting Preprint arXiv:1709.04875

[3] Yao H, Liu Y, Wei Y, Tang X and Li Z (2019) Learning from multiple cities: A meta-learning approach for spatial-temporal prediction Proc. The World Wide Web Conf. pp 2181-2191. https://doi.org/10.1145/3308558.3313577

[4] Tsapakis I, Cheng T and Bolbol A (2013) Impact of weather conditions on macroscopic urban travel times J. Transp. Geogr. 28 204-211. https://doi.org/10.1016/j.jtrangeo.2012.11.003

[5] Hranac R, Sterzin E, Krechmer D, Rakha H A and Farzaneh M (2006) Empirical studies on traffic flow in inclement weather

[6] Xu F, He Z, Sha Z, Zhuang L and Sun W (2013) Assessing the impact of rainfall on traffic operation of urban road network Procedia - Social and Behavioral Sciences 96 82-89. https://doi.org/10.1016/j.sbspro.2013.08.012

[7] Zhong H, Wang J, Chen C, Wang J, Li D and Guo K (2024) Weather Interaction-Aware Spatio-Temporal Attention Networks for Urban Traffic Flow Prediction Buildings 14 647. https://doi.org/10.3390/buildings14030647

[8] Cools M, Moons E and Wets G (2010) Assessing the impact of public holidays on travel time expenditure: differentiation by trip motive Transp. Res. Rec. 2157 29-37. https://doi.org/10.3141/2157-04

[9] Yu W, Zhao D, Hua X, Wen H, Lei H and Wang W (2024) Spatiotemporal dynamics and determining factors of intercity mobility: A comparison between holidays and non-holidays in China Cities 153 105306. https://doi.org/10.1016/j.cities.2024.105306

[10] Chen N, Li B, Tao J and Jin S (2021) A method for daily traffic flow parameter forecasting combining the impact of holidays Int. Conf. on Intelligent Transportation Engineering (Singapore: Springer Nature Singapore) pp 704-715. https://doi.org/10.1007/978-981-19-2259-6_63

[11] Jin Y, Chen K and Yang Q (2023) Transferable graph structure learning for graph-based traffic forecasting across cities Proc. 29th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining pp 1032-1043. https://doi.org/10.1145/3580305.3599529

[12] Shao Z, Zhang Z, Wang F and Xu Y (2022) Pre-training enhanced spatial-temporal graph neural network for multivariate time series forecasting Proc. 28th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining pp 1567-1577 https://doi.org/10.1145/3534678.3539396

[13] Shang C, Chen J and Bi J (2021) Discrete graph structure learning for forecasting multiple time series Int. Conf. on Learning Representations (ICLR)

[14] Wu Z, Pan S, Long G, Jiang J and Zhang C (2019) Graph wavenet for deep spatial-temporal graph modeling Proc. 28th Int. Joint Conf. on Artificial Intelligence pp 1907-1913

[15] Li Y, Yu R, Shahabi C and Liu Y (2018) Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting Proc. 6th Int. Conf. on Learning Representations

[16] Song C, Lin Y, Guo S and Wan H (2020) Spatial-Temporal Synchronous Graph Convolutional Networks: A New Framework for Spatial-Temporal Network Data Forecasting Proc. 34th AAAI Conf. on Artificial Intelligence pp 914-921

[17] Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, Deng M and Li H (2019) T-GCN: A temporal graph convolutional network for traffic prediction IEEE Trans. Intell. Transp. Syst. 21 3848-3858