ITM Web of Conferences **70**, 01004 (2025) https://doi.org/10.1051/itmconf/20257001004
DAI 2024

Application Research of Cross-Attention Mechanism for Traffic Prediction Based on Heterogeneous Data

Zhihao Feng

Faculty of Data Science, City University of Macau, Macau, 999078, China

Abstract. Intelligent transportation systems need to be developed with precise traffic flow predictions to reduce traffic accidents, improve overall urban mobility, and mitigate congestion. The intricacy and variety of traffic conditions are often too complex and variable for traditional approaches to handling, especially when dealing with heterogeneous event data from several sources like weather variations and traffic incidents. This review highlights the significance of cross-attention mechanisms by examining the developments in integrating multi-source heterogeneous event data for traffic prediction. Examining different approaches used in previous work, the study focuses on the Event-aware Graph Attention Fusion Network (EGAF-Net). This cutting-edge model efficiently integrates and analyzes complex spatial-temporal data. Through an analysis of these methods, the research demonstrates how applying advanced deep learning algorithms and cross-attention processes has significantly improved prediction robustness and accuracy. The results underscore the critical role of heterogeneous data integration in enhancing traffic prediction models, providing insights into current challenges and potential future developments in the field. This thorough analysis aims to direct future research endeavors and open the door for more dependable and effective intelligent transportation systems.

1 Introduction

Predicting traffic flow is a crucial component of intelligent transportation systems. Traffic congestion not only affects travel efficiency but also may lead to a series of problems such as traffic accidents, environmental pollution, and economic losses. Predicting traffic movement is therefore especially crucial. It is an important method to achieve urban traffic optimization in intelligent transportation systems. For traffic management and guiding, accurate traffic flow forecast is crucial [1]. By predicting future traffic conditions, it can effectively improve traffic management, reduce congestion, and improve road utilization efficiency. There are numerous ways to estimate traffic flow, but they can be broadly categorized into three categories: classic time series prediction, standard machine learning, and deep learning. Although these methods can provide useful predictions in some cases, they usually fail to capture emergencies in traffic flow and their impact.

Corresponding author: D22090101344@cityu.edu.mo

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

However, with the frequent occurrence of traffic incidents, prediction methods that rely solely on historical data are difficult to reflect real-time traffic conditions. These traditional methods show obvious limitations when facing emergencies, such as the inability to update data models promptly and the lack of response capabilities to emergencies, resulting in reduced prediction accuracy. Event data refers to various dynamically changing information in the traffic system, which can be used to predict traffic flow conditions. Event data includes traffic accidents, road construction, weather changes, etc., and can provide more real-time and dynamic information than historical traffic data.

Liu et al. designed a method called learning vector quantization (LVQ) network DFMIAID to address the issue of sparse numbers and widely spaced detectors. This data fusion method can not only make full use of existing information resources, but also does not require additional information investment, and has good development prospects [2]. Cai et al. studied a highway traffic incident detection method called DFMIAID, which is based on support vector machine (SVM) and data fusion principles. The research results show that among the three SVM models, the approximately linearly separable support vector machine has a relatively better DFMIAID event detection effect. The use of these techniques is anticipated to greatly increase the precision and effectiveness of traffic flow forecast and generate fresh concepts for resolving issues with traffic congestion [3]. To increase the accuracy of traffic flow prediction, the research methodology used in this work primarily consists of fusing multi-source heterogeneous event data utilizing a deep learning model and cross-attention mechanism. To enhance the effectiveness of the prediction model, the study topic focuses on how to combine and handle data from many sources, including weather variations and traffic incidents. The goal of the project is to improve traffic management efficiency by strengthening the traffic flow prediction model and making it more accurate and real-time through the use of efficient data fusion technologies.

2 Method overview

2.1 Event data

Vehicle video detection technology has the advantages of high precision, real-time monitoring, diversity, and flexibility in highway systems. It can accurately identify vehicle types and speeds, detect traffic accidents, violations, road congestion, and other events, and can be combined with other detection technologies such as microwave vehicle detectors to improve the overall detection effect [4]. It can be seen that event data usually has the characteristics of suddenness, timeliness, and diversity, and can provide more real-time and dynamic information than historical traffic data.

2.2 Traffic flow prediction method

The sudden nonlinearity that results from the changes between free flow, breakdown, recovery, and congestion makes it difficult to forecast traffic flow [5]. The three primary parts of the traffic accident prediction model are as follows: the initial part is the spatial learning layer, which uses graph convolution operations to learn the correlation in spatial information; the next is the spatiotemporal learning layer, which combines standard convolution and graph convolution to capture dynamic changes in spatial and temporal perspectives; and the final part is the embedding layer, which works to extract meaningful semantic representations of external information [6].

2.3 Traditional methods

Traditional statistical models such as time series analysis (ARIMA model and its variants) and regression models are generally used when traffic flow data is relatively stable. Nevertheless, the nonlinear and dynamic changes in traffic flow are not well captured by these models, which are predicated on the assumption of linear and stable time series.

2.4 Machine learning methods

Machine learning models include support vector machines and random forests, which can handle nonlinear relationships and high-dimensional data. In traffic flow prediction, machine learning models make real-time predictions by learning the relationship between historical data and event data, but their overall nonlinear processing capabilities are limited and are not optimal for modeling complex and dynamic traffic data.

2.5 Deep learning methods

Convolutional neural networks, recurrent neural networks, and graph neural networks are the primary neural network types used in deep learning models. These networks process image, time series, and graph data, respectively, and automatically extract features through hierarchical structures that can capture intricate dynamic changes in traffic. In addition, there are some innovative models such as grid models, graph models, STA-RNIMEI models, etc. The grid model combines external feature information, extracts spatiotemporal features through convolution operations and residual networks, and finally uses visual converters to capture global features for prediction. The graph structure model uses graph neural networks and self-attention mechanisms, combined with graph convolution modules to model complex spatiotemporal dependencies, enhance feature representation and decoder output, and raise the traffic predictions accuracy. The STA-RNIMEI model combines a variety of external information for modeling, captures the temporal dynamics of POI data through time perception and channel attention mechanisms, and models meteorological data with fully connected networks and convolutional networks [7, 8]. Federated learning updates the learning model through a secure parameter aggregation mechanism without directly sharing local data sets. It is based on the long short-term memory network (LSTM) federated learning algorithm and decentralized federated learning traffic flow prediction based on blockchain technology. Decentralized federated learning based on blockchain technology implements a decentralized federated learning solution without a central server and ensures the reliability and security of model update data through miner verification and consensus mechanism on the blockchain [9].

The development of these three generations of methods reflects the researchers' deepening understanding of traffic flow prediction, from simple linear prediction to complex spatiotemporal series modeling, and the gradual increase in data samples (from early fixed-point observations to modern global mobile trajectory data). In addition, the modeling capabilities of the prediction models have also been continuously enhanced, from linear hypothesis models in the statistical field to shallow machine learning models that can handle nonlinear and high-dimensional data, to deep neural networks with huge parameter scales. These changes have promoted the continuous advancement of traffic flow prediction technology, making it more accurate and efficient when processing complex and dynamic traffic flow data [10].

3 Practical application cases

3.1 Traffic accident prediction

In areas where traffic accidents frequently occur, the introduction of event data combined with historical traffic flow data can more quickly predict the impact of accidents on traffic flow, Consequently, early guidance is given and traffic congestion is reduced. Liu suggested an ensemble learning-based approach for predicting the severity of traffic accidents based on the features of various models' applications. It fully considers time, space, and other characteristic data that may affect the occurrence of accidents, and uses the advantages of the combined model to deeply mine regional road traffic accident data, thereby improving the prediction accuracy of accident severity [10].

3.2 Weather impact forecast

Numerous studies have examined how weather conditions affect traffic accidents [11, 12]. From this, it can see the impact of climate factors on traffic accidents. Weather data includes slippery, snowy, or icy roads, muddy, etc [13]. Traffic management departments can take appropriate countermeasures by anticipating the influence of weather variations on traffic flow and using real-time weather information to build a traffic flow prediction model. By conducting a thorough analysis of traffic flow characteristics, researchers have produced useful tools for comprehending the dynamic features of traffic flow under various circumstances. They have also produced substantial foundational information and more comprehensive concepts for developing a short-term weather-related traffic flow prediction model. Based on these research findings, Nico Becker further examined the periodic characteristics, long-term temporal correlation, spatial correlation, and the mechanism by which weather conditions affect traffic flow in the road network. The research then built a model based on a Deep Learning short-term traffic flow prediction model using the attention mechanism, graph neural network and gated recurrent unit. The model's example verification findings using large-scale real road network traffic flow data demonstrate a considerable improvement in prediction performance when compared to previous models, indicating the model's high prediction accuracy and practical usefulness [13].

3.3 Traffic management for large events

Concerts and athletic events are examples of planned special events (PSEs) that might result in non-recurring traffic jams. This is because large numbers of attendees leave the venue at the same time, overloading the local transportation network and causing serious traffic problems [14]. Event data can be used to predict changes during events, helping traffic management departments optimize traffic plans and ensure smooth traffic. Pan et al. analyzed passenger flow data during large-scale events and discussed the data sets of dispersed continuous and concentrated gathering and dispersion events respectively. For dispersed continuous events, a prediction model for inbound and outbound passenger flow was constructed, CNN was used to select features, LSTM and BiLSTM+Attention models were used for prediction, and the results were verified by actual event data. For concentrated gathering and dispersion events, the researchers analyzed their characteristics, predicted passenger flow in different impact periods, and linked the outbound and inbound volumes of urban rail transit. Through case analysis and comparative experiments, the feasibility of the model and the auxiliary role of multi-venue information in subway passenger flow prediction were verified [15].

4 Traffic prediction and multi-source heterogeneous event knowledge fusion

The goal of multimodal visualization is to integrate various data sets to provide users with more comprehensive information and deeper understanding [16]. Multimodal big data is made up of several modalities, each of which has an independent distribution and offers a partial description of the same object of interest. There are intricate relationships between these modes, though. The goal of multimodal data fusion is to uniformly represent this data, which come from many sources, distributions, and types, in a single space. It is possible to enhance the performance of different multimodal applications by modeling cross-modality and multimodality [17]. A common method now is to directly concatenate the feature vectors obtained by encoding each event data. Although this approach is simple to use, it might not completely take advantage of the relationships that could exist between different kinds of data. The simply concatenated feature vectors lack interactivity and cannot dynamically adjust the importance of different modal features, which may lead to poor information fusion and reduce prediction accuracy. Therefore, a fully connected neural network is used on this basis to further process and map features. This approach can improve the model's capacity for nonlinear expression, but it still has to be tuned in terms of network structure design and hyperparameters. Therefore, there is also a multi-task learning method, which trains multiple related tasks simultaneously to share and optimize feature representations. This strategy can increase the model's capacity for generalization, but the design of the trade-offs between tasks and sharing methods is quite difficult. In order to accurately represent the spatiotemporal relationships and event effects in the prediction of traffic speed, Qiu created an event-aware graph attention fusion network (EGAF-Net) [18]. The spatial event embedding block and the dynamic event graph constructor learn the spatial and temporal representation of events, the gated fusion mechanism integrates the spatiotemporal correlation in the road network with the event representation, and the ST-Speed attention block simulates the spatial correlation of road segments [18].

When studying multi-source heterogeneous events, the Transformer's cross-attention mechanism is widely used to match and associate multimodal information. Through neural network learning, it can efficiently capture long-distance dependencies in sequence data and accomplish data-driven end-to-end multimodal information fusion [19]. The CoCa (Contrastive Captioner) model integrates the advantages of single encoder, dual encoder, and encoder-decoder paradigms for processing image and text data. In its model architecture, a cross-attention mechanism is applied between the image encoder and the multimodal decoder to learn multimodal image-text representations. This architecture uses cross-attention to enhance the fusion effect of multimodal information by decoupling the unimodal decoder and the multimodal decoder. This method demonstrates the effectiveness of cross-attention in processing multimodal data [20].

5 Conclusion

In addition to introducing the study methodologies and applications of multi-source heterogeneous event knowledge fusion and traffic flow prediction, this paper also examines the use of conventional statistics, machine learning, and deep learning techniques in this field. In conclusion, it is discovered that the accuracy and real-time performance of traffic flow prediction can be enhanced by the addition of event data, such as weather variations and traffic incidents. Furthermore, the present state of multi-source heterogeneous data fusion is examined, and a deep learning model-based fusion technique is suggested. The accuracy and resilience of the traffic flow prediction model can be increased by fusing data from multiple sources of heterogeneous events. Event data provides real-time and dynamic information, which enables the prediction model to better respond to emergencies and thus provide reliable traffic flow predictions. This article findings support earlier research showing that event data is crucial for predicting traffic flow. However, this study further verifies the effectiveness of multi-source heterogeneous data fusion by introducing more types of event data, such as weather changes and large-scale events. These findings provide new approaches to existing theories and practices and further demonstrate the potential of multi-source data fusion technology in traffic flow prediction. This study is subject to some constraints, including the possibility that the model's accuracy could be impacted by the variety and caliber of data sources. Furthermore, the computational cost of the model rises due to its complexity, which could restrict its real-time performance in real-world applications.

To further enhance the prediction performance, future studies should concentrate on data control and model optimization. The following directions for future research can be taken into consideration based on the findings of this paper: Initially, investigate the multi-source data fusion method further to enhance the emergency response capability of the model. Second, real-time data processing technologies can improve the model's efficacy in real-time. In conclusion, additional forms of event data, like information from traffic sensors and social media platforms, can be integrated to enhance the prediction model's sources. Overall, this study confirms the efficacy of multi-source data fusion technology in traffic flow prediction, offers a thorough understanding of the fusion of traffic flow prediction and multi-source heterogeneous event knowledge, and lays a strong platform for further research and application. Traffic flow prediction models can more effectively handle the effects of emergencies and produce more accurate and dependable prediction results by combining event data. Finally, it is hoped that this study can make certain contributions to the development of related fields and trigger more meaningful discussions. Traffic flow prediction technology will develop further via study and real-world applications. It will also strongly support the creation of intelligent transportation systems, which will enhance traffic control, boost travel efficiency, and encourage sustainable social and economic growth.

References

1. J. M. Hu, Research on short-term traffic flow prediction method under the influence of weather factors, Master's thesis, Jilin Univ. (2021).
2. W. M. Liu, X. Y. Yin, L. P. Guan, An event detection algorithm based on multi-information fusion, J. Changsha Commun. Univ. (01), 58-62 (2004).
3. Z. Li, G. J. Guiyan, Y. L. Yong, Algorithms of SVM-AID based on data-level and decision-making-level data fusion methods, in 2008 Chinese Control and Decision Conference, IEEE, 3851-3856 (2008).
4. Y. Y. Chen, Z. L. Pan, Q. S. Wu, et al., Research on traffic event detection system based on multi-data fusion, Highway Traffic Technol. (App. Technol. Ed.) **8**(08), 375-377 (2012).
5. N. G. Polson, V. O. Sokolov, Deep learning for short-term traffic flow prediction, Transp. Res. Part C: Emerg. Technol. **79**, 1-17 (2017).
6. L. Yu, B. Du, X. Hu, et al., Deep spatio-temporal graph convolutional network for traffic accident prediction, Neurocomputing **423**, 135-147 (2021).
7. X. H. Sun, Analysis and prediction of urban traffic flow based on deep learning, Master's thesis, Jilin Univ. (2023).
8. Y. M. Dong, Traffic flow prediction based on spatio-temporal feature learning, Master's thesis, Univ. of Electronic Sci. and Technol. China (2022).

9. H. M. Tang, Research on privacy-preserving traffic flow prediction scheme based on federated learning, Master's thesis, East China Normal Univ. (2023).
10. Y. H. Liu, H. X. Shen, Spatio-temporal prediction of traffic accident severity based on ensemble learning models, Sci. Technol. Innov. Appl. **14**(08), 28-35 (2024).
11. B. Zhai, J. Lu, Y. Wang, et al., Real-time prediction of crash risk on freeways under fog conditions, Int. J. Transp. Sci. Technol. **9**(4), 287-298 (2020).
12. F. Malin, I. Norros, S. Innamaa, Accident risk of road and weather conditions on different road types, Accid. Anal. Prev. **122**, 181-188 (2019).
13. N. Becker, H. W. Rust, U. Ulbrich, Predictive modeling of hourly probabilities for weather-related road accidents, Nat. Hazards Earth Syst. Sci. **20**(10), 2857-2871 (2020).
14. S. Lassacher, D. Veneziano, S. Albert, et al., Traffic management of special events in small communities, Transp. Res. Rec. **2099**(1), 85-93 (2009).
15. W. T. Pan, Short-term passenger flow prediction of urban rail transit under large-scale events, Master's thesis, Beijing Jiaotong Univ. (2023).
16. R. Bramon, I. Boada, A. Bardera, et al., Multimodal data fusion based on mutual information, IEEE Trans. Vis. Comput. Graph. **18**(9), 1574-1587 (2011).
17. J. Gao, P. Li, Z. Chen, et al., A survey on deep learning for multimodal data fusion, Neural Comput. **32**(5), 829-864 (2020).
18. Z. Qiu, T. Zhu, Y. Jin, et al., A graph attention fusion network for event-driven traffic speed prediction, Inf. Sci. **622**, 405-423 (2023).
19. B. L. Zhang, Z. H. Pan, J. Z. Jiang, et al., Multi-modal perception fusion method based on cross-attention mechanism, China J. Highway Transp. **37**(03), 181-193 (2024).
20. J. Yu, Z. Wang, V. Vasudevan, et al., Coca: Contrastive captioners are image-text foundation models, arXiv preprint arXiv:2205.01917 (2022).