The Application of Reinforcement Learning in Traffic Flow Prediction: Advantages, Problems, and Prospects

Minghui Li[1*], *Decheng* Zhou[2], and *Shiqi* Zhang[3]

[1]SWUFE-UD Institute of Data Science at SWUFE, Southwestern University of Finance and Economics, Chengdu, Sichuan, 611130, China
[2]School of Information Science and Technology, ShanghaiTech University, Shanghai, 200120, China
[3]Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, 48109, USA

Abstract. Traffic flow prediction (TFP) is an important topic in the fields of operation research and traffic engineering. It is dedicated to predicting the flow of people and vehicles in the transportation network within a specific time frame in the future. Accurate TFP has great significance for traffic management, urban planning, road design, and the development of intelligent transportation systems (ITS). This article summarizes three traditional methods of TFP: parameter-based prediction, shallow machine learning-based prediction, and deep learning (DL)-based prediction. However, traditional TFP methods only focus on predicting time series in traffic data, and it is difficult for these methods to capture the interdependent relationship between the spatial distribution of traffic across a network and the temporal evolution of traffic conditions at each location. sequences. How to fully extract the spatiotemporal correlation of traffic flow (SCTF) is an urgent problem that needs to be solved based on DL prediction models. Concurrently, as science and technology advance, a growing variety of academics are attempting to incorporate reinforcement learning (RL) into TFP. Experimental results show that it can reduce vehicle queuing time and average delay to a greater extent, and alleviate air pollution. The article summarizes the models of DL and RL in TFP, comprehensively compares the benefits and drawbacks of various approaches, and proposes a vision for existing problems and future development.

1 Introduction

With the acceleration of urbanization and the sudden surge in the number of motor vehicles, China's traffic congestion (TC) problem requires immediate resolution. According to the "2023 China Urban Transportation Report", 86% of China's 100 cities in 2023 have urban commuting peak TC index (an indicator that characterizes urban TC, that is, the actual travel time and smooth traffic during morning and evening peak hours on workdays)

* Corresponding author: jimmylee@udel.edu

increased year-on-year in 2022, with an average increase of 7.17% and a maximum increase of 26.92% [1]. This shows that as the public's enthusiasm for travel increases, some cities have experienced a significant increase in TC. By collecting traffic flow data and using ITS to predict and optimize traffic flow, improving traffic efficiency and reducing negative impacts has become a way to alleviate TC.

There are currently three relatively mature traditional forecasting methods in the field of traffic flow forecasting. The first is the autoregressive integrated moving average method (ARIMA), which is a versatile and popular technique for predicting a range of time series data; the second is the well-known shallow ML method. Its model usually has only one or several hidden layers. Compared with the DL model, it has fewer parameters and is fast in training, but it is not suitable for complex problems. The fitting ability is weak; the third type is the DL method, which is a branch of ML and can automatically learn high-level features of data through multi-layer neural networks and has strong fitting ability for complex problems. RL is an important branch of ML. Its unique advantage lies in its ability to automatically learn and optimize decision-making strategies without explicit guidance through continuous interaction between the agent and the environment to maximize cumulative rewards. Traditional TFP methods have limitations in dealing with real-time and dynamic characteristics. The RL method provides new ideas for real-time TFP with its excellent adaptability and learning ability.

RL-based TFP methods have been successful in recent years. Deep reinforcement learning (DRL) has resulted in a high prediction accuracy of 'traffic flow' and 'average speed' for the long short-term memory (LSTM) model [2]. TFP values of the RL method model based on Recurrent Neural Network (RNN) have a small error compared to the actual observed values [3]. The traffic flow optimization prediction method based on clustering and RL attracts a mean absolute percentage error that falls below 3.25%, leading to high prediction accuracy [4].

The purpose of this article is to discuss the technique for optimizing traffic flow using RL. By comparing the effects of different model methods in TFP, it is concluded that the TFP method based on RL is conducive to improving the accuracy of TFP and is helpful to Reducing TC is of great significance to improving urban traffic conditions.

2 Methods

Traditional TFP models can be divided into the following three types according to model prediction methods: the first one is prediction based on parameter, prediction based on shallow machine learning, and prediction based on DL.

The most typical representative of prediction methods based on parameter is the ARIMA model method. The ARIMA model method is a statistical method used for time series analysis and forecasting. The ARIMA model captures different patterns in time series by combining three parts: autoregression, integration, and moving average. However, although the ARIMA model has good performance in the TFP time series, it is difficult to capture the SCTF sequences.

Prediction techniques based on shallow ML primarily incorporate the KNN algorithm and support vector regression (SVR). SVR is a supervised learning algorithm that is extended from the support vector machine (SVM) and is used to solve regression problems. Unlike SVM, which determines sample categories in classification problems, SVR aims to find a function that makes the predicted value as close as possible to the actual value while keeping the complexity of the model as low as possible. The KNN algorithm is a simple and intuitive supervised learning algorithm that is widely used in classification and regression problems. The KNN algorithm finds the most comparable k neighbors by computing the distance between a new data point and every data point in the training data

set. It then predicts the classification or regression value of a new data point, assuming that similar occurrences have similar outputs.

Prediction methods based on DL mainly include LSTM, gated recurrent unit neural network (GRU) and stacked auto-encoder neural network (stacked auto-encoders, SAEs), etc. LSTM is a special type of RNN, which is designed to solve the gradient vanishing and gradient explosion problems of standard RNN when dealing with long-term dependence problems. LSTM can better capture and store the long-term dependencies of time series data by introducing a series of gating mechanisms to control the flow of information. GRU is an improved RNN architecture proposed by Cho et al. in 2014 [5]. The design of GRU is inspired by the LSTM network, but compared to LSTM, GRU has a simpler structure and performs more efficiently on certain tasks. Stacked Autoencoder (SAE) is a deep neural network architecture. Autoencoder is an unsupervised learning model whose purpose is to learn a way to efficiently encode and decode input data so that the reconstructed output is as close as possible to the original input. SAEs train multiple autoencoders layer by layer to extract high level features of data layer by layer for tasks such as feature extraction, and data denoising. With the enhancement of computing power and the rapid development of artificial intelligence, DL models are frequently employed in the field of TFP and have emerged as a major research focus. However, prediction methods based on DL also have some problems. Traditional VAR models, ARIMA models, and SVR models only focus on predicting time series in traffic data, and these methods do not pay attention to the spatiotemporal correlation between different locations. The problem of how to fully extract the characteristics of the SCTF is an issue that needs to be urgently solved based on DL prediction models. Table 1 summarizes the limitations of three different types of forecasting methods.

Table 1. The limitations of three different types of forecasting methods.

Paper	Model & Method	Limitation
[6]	Model: ARIMA Method: Parameter-based forecasting method	The SCTF sequences are challenging to capture.
[7]	Model: SVR Method: Prediction method based on shallow ML	When the SVR prediction model encounters high dimensional data, the processing speed is relatively slow and the calculation is expensive. The cost is relatively high and there is a delay in the output.
[8]	Model: LSTM Method: Prediction methods based on DL	The accuracy of the predictions will be impacted since the spatial information needs to be manually entered as the network's input due to the spatial peculiarities of the data.
[9]	Model: CNN - LSTM Method: Prediction methods based on DL	When the traffic network topology changes, the DL model may need to be retrained. Adaptability to dynamic environments may be insufficient: Compared with the good performance of DL models in static or slowly changing traffic environments, highly dynamic and continuously changing traffic scenarios may weaken their actual performance.
[10]	Model: GNN-CL Method: Prediction methods based on DL	Lack of transferability: Traffic flow patterns may vary greatly in different areas, and the results obtained by training the model in one area may not transfer well to another area. This means that if a Traffic Stream model performs well in one city, it may not necessarily work equally well in another city.

At the same time, in the development process of TFP, due to the problem that DL methods cannot accurately extract the SCTF, some scholars have proposed using RL in TFP. RL is a type of ML that involves how to let an agent take actions in the environment to maximize the cumulative return. This method is widely used in a variety of problems, from strategy selection in games to robot control to adaptive systems. RL algorithms can continuously update and improve their own strategies based on feedback from the environment. For TFP, this means that the model can continue to learn from new data and continuously improve prediction accuracy. There are often complex nonlinear relationships between various factors in traffic flow (such as traffic volume, traffic lights, weather, etc. Traditional linear or static methods may have difficulty capturing these relationships, while RL, especially methods combined with DL techniques, has powerful nonlinear function approximation capabilities and can better handle these complex relationships. Chen et al. proposed a LSTM model based on DRL to predict traffic flow [2]. The developed model that combines DRL and LSTM can reliably forecast traffic flow data, according to experimental results. In addition, road traffic conditions can be accurately reflected by the traffic condition representation model. Of them, the "traffic flow" forecast accuracy is 97.10%, the "average speed" prediction accuracy is 99.34%, and the "traffic conditions" prediction accuracy is 85.71% [2]. However, in the prediction results using only the LSTM model, the segment prediction accuracy rates of "traffic flow", "average speed" and traffic conditions are 94.42%, 99.07% and 81.40% respectively [2]. This shows that applying RL algorithms to TFP will effectively improve the accuracy of TFP. Zhao et al. proposed a TFP method based on RL and verified this method using the METR-LA dataset [3]. Experimental results show that this method has relatively good prediction results in different scenarios: root mean square error is 1.8 to 5.2, and average absolute error is 1.2 to 3.8, both at low levels [3]. This method is relatively good during stationary periods, but the prediction accuracy during peak periods needs to be improved.

3 Problems and challenges

3.1 Sparse reward problem

Conventional RL with a reward signal is essential in guiding the agent to learn a good policy. However, in a scenario aiming for the prediction of traffic flow, the actions the agent takes, such as dealing with the traffic signal control or recommending the route, take a long time for the effect to be realized. Thus, the reward signal is delayed and sparse, only given after several steps. This scarcity thus makes it hard for a RL algorithm to learn quickly and associate its actions directly with the outcomes of those actions just from the reward signal. Haarnoja et al. have observed that sparse rewards can dramatically increase the length of time it takes for a RL algorithm to explore the optimal strategy during training. Although their research involved the introduction of Deep Energy-Based Policies to tackle the problem, the result has still been that such approaches require considerable computational resources and their practical application remains limited [11].

3.2 High-dimensional state space problem

In the case of TFP, there exists involvement with several variables, including traffic density, vehicle speed, road conditions, weather, and unforeseen events, which lead to very high-dimensional state spaces. Thus, this high dimensionality of the state space gives rise to the "curse of dimensionality": how the size of the sample required grows exponents in size with every new dimension added, reducing the learning efficiency of RL algorithms.

Furthermore, this complicates the task of the agent in fully exploring this state space within the time constraints usually returned for training, usually returning suboptimal strategies. Li et al. designed a Graph-based RL method for traffic signal control problems, noting that, in TFP, the high-dimensional state space significantly raises the computational cost and time for training RL models. Even though they use the structure of graphs to curtail some of these issues, optimization of the models in high-dimensional settings is still a daunting task [12].

3.3 Uncertainty in traffic flow data

Traffic flow data is usually highly variable and uncertain due to weather changes, sudden accidents, road repairs, holiday effects, and even random driver behavior. This makes the prediction of traffic flow highly complex and puts higher demands on the robustness of RL models. It is thus important that RL algorithms be greatly robust to these unstable learning curves during training, where there could be big swings in the predictive performance of a model. Furthermore, since these models have uncertainty, rare but severe traffic incidents can hardly be handled in real-life applications. Based on these, Zheng et al. showed that the uncertainty in the traffic flow has been a remarkable barrier to RL algorithms, mostly in the situation of dynamic traffic control. Although it added an RL-based dynamic control parameter tuning approach, even then it was not working very well under highly uncertain traffic data and even worse during events [13].

3.4 Model interpretability problem

The credibility and acceptance of TFP models when it is used in real-world applications is a function of the interpretability of the results. The outcome of the decision concerning traffic management has direct consequences on the safety and welfare of the citizens. Thus, stakeholders such as traffic managers or policymakers often require models to be transparent about the reasoning behind predictions or decisions. Most DL-based RL models, however, are normally "black-box" models, whose complex internal structures obscure their processes of decision-making. The lack of transparency of the model may therefore hamper the broader applicability of RL in high-stake areas. Loyola-González presented the pros and cons of black-box and white-box models, noting that although black-box models, as in deep RL, often equally yield useful prediction performances, the low interpretability makes it hard for a decision-maker to confidently apply those in very important fields such as traffic management. This thus calls for a trade-off between model performance and interpretability to make more applicability to real-world problems possible [14].

3.5 Real-time requirement

Traffic flow data is usually highly variable and uncertain due to weather changes, sudden accidents, road repairs, holiday effects, and even random driver behavior. This makes the prediction of traffic flow highly complex and puts higher demands on the robustness of RL models. It is thus important that RL algorithms be greatly robust to these unstable learning curves during training, where there could be big swings in the predictive performance of a model. Furthermore, since these models have uncertainty, rare but severe traffic incidents can hardly be handled in real-life applications. Based on these, Zheng et al. showed that the uncertainty in the traffic flow has been a remarkable barrier to RL algorithms, mostly in the situation of dynamic traffic control. Although it added an RL-based dynamic control parameter tuning approach, even then it was not working very well under highly uncertain traffic data and even worse during events [13].

3.6 Model interpretability problem

The credibility and acceptance of TFP models when it is used in real-world applications is a function of the interpretability of the results. The outcome of the decision concerning traffic management has direct consequences on the safety and welfare of the citizens. Thus, stakeholders such as traffic managers or policymakers often require models to be transparent about the reasoning behind predictions or decisions. Most DL-based RL models, however, are normally "black-box" models, whose complex internal structures obscure their processes of decision-making. The lack of transparency of the model may therefore hamper the broader applicability of RL in high-stake areas. Loyola-González presented the pros and cons of black-box and white-box models, noting that although black-box models, as in deep RL, often equally yield useful prediction performances, the low interpretability makes it hard for a decision-maker to confidently apply those in very important fields such as traffic management. This thus calls for a trade-off between model performance and interpretability to make more applicability to real-world problems possible [14].

4 Future research directions

4.1 Addressing the sparse reward problem

However, the sparse reward problem remains one of the most urgent and unsolved issues in RL, especially in the context of TFP. In this sense, the future line of research might orient one way to the development of stronger strategies for reward, either through intermediate rewards or surrogate models that allow for much more frequent feedback. In addition, it is possible to use such methods as imitation learning or inverse RL for more exact fine-tuning of strategy-building efficiency and quality in a problem of having sparse rewards.

4.2 Handling high-dimensional state spaces

The high-dimensional state space in TFP, to date, remains one of the major contributing factors that increase the complexity of RL algorithms. Future research in this area may aim at the development of more advanced dimensionality reduction with techniques such as Graph Neural Networks or Autoencoders. More importantly, the incorporation of attention mechanisms can help pick out and focus only on critical features that will enhance the efficiency of RL algorithms.

4.3 Improving model robustness to address uncertainty

One of the major challenges of RL algorithms is the inherent uncertainty in traffic flow data. Some other promising strategies in future studies would be methods that enhance model robustness under uncertain conditions, such as Bayesian RL or Distributional RL. Besides, online learning methods make models dynamically adjust themselves to changes in real-time data and enhance resilience against uncertainty.

4.4 Improving model interpretability

One long-term priority in improving the interpretability of RL models would be to develop more transparent models of decisions or consider hybrid architectures that join interpretable, rule-based methods with techniques from DL. Moreover, visualization tools and frameworks of interpretability could be leveraged to shed more light on decision-making

processes for the more complex models developed by RL, making them more accessible and trustworthy for various stakeholders.

4.5 Improving sample efficiency and training speed

To address this rather poor sample efficiency, possible future research directions could involve Transfer Learning or Meta-Learning methods that possibly require less extensive training datasets. This could also be achieved by further optimization of training methodologies and parallel processing techniques in use. Another source of training data is simulation environments or synthetic data generated by Generative Adversarial Networks, which could add to the sample efficiency.

4.6 Real-time requirements

In the future, solving the real-time requirements for the prediction of traffic flow would necessitate research on faster RL low-latency online learning methods. Research on models that apply stream data processing, able to dynamically update the strategies in real-time, can greatly improve the real-time prediction capacity. Furthermore, the deployment of RL models in distributed computing environments and making full use of multi-node parallel processing will dramatically improve the real-time responsiveness of the models.

4.7 Optimizing task weights and interactions in MTL

Future research on MTL will implement self-adjustment of task weights and further design more sophisticated mechanisms for interactions across tasks that can aid in better balancing of allocations of resources between different tasks. Innovative MTL architectures can manage these conflicts and synergies to further arise between the tasks, therefore improving the overall model performance and stability.

5 Conclusion

This paper concludes that the approaches to predict traffic flow using RL can do better than the three traditional methods in prediction accuracy and handling ability to pass the dynamic and complex features of traffic flow data. Challenges remain regarding dealing with high-dimensional state spaces and managing uncertainty inherent in traffic flow data while improving real-time performance with such models. Much more research on this topic is needed to optimize further the RL algorithms with efficient dimensionality reduction techniques and proper handling of uncertainty measures. In addition, it will incorporate MTL and Transfer Learning techniques to improve generalization performance and adaptability so that the models can handle the increasingly complex and variable urban traffic environments that are pushing the technologies in flow prediction into new levels.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

1. Baidu Maps. 2023 China urban transportation report. Baidu, March 1, 2024. Retrieved on August 13, 2024. Available at: https://jiaotong.baidu.com/cms/reports/traffic/2023/index.html.
2. Z. Chen, X. Luo, T. Wang, W. Wang, & W. Zhao. Deep reinforcement learning-based lstm model for traffic flow forecasting in internet of vehicles. In Z. Deng (Ed.), Proc. 2021 Chinese Intelligent Automation Conf. Lect. Notes Electr. Eng., **801**, Springer, Singapore (2022).
3. Y. Zhao, & Y. B. Zhou. Application research of reinforcement learning in traffic flow prediction. Inf. Comput. (Theor. Ed.), **36**(03), 136-138 (2024).
4. S. C. Rajkumar, J. Deborah L., & P. Vijayakumar. optimized traffic flow prediction based on cluster formation and reinforcement learning. Int. J. Commun. Syst. **33**, e4178 (2019).
5. K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, & Y. Bengio. Learning phrase representations using RNN Encoder–Decoder for statistical machine translation. In Proc. 2014 Conf. Empirical Methods Nat. Lang. Process. (EMNLP), 1724-1734, Doha, Qatar. Assoc. Comput. Linguist. (2014).
6. M. M. Tan, X. G. Cheng, & K. Zhou. Short-term traffic flow prediction based on weighted combination of arima and grey model. Comput. Technol. Dev., **26**(11), 77-81 (2016).
7. Y. Liu, X. Li, & X. Shao. Short-term traffic flow prediction based on lagrange support vector regression. Comput. Commun., 1249-1254 (2007).
8. B. Cao, & M. T. Gao. Research on the short-term traffic flow prediction based on LSTM. Mod. Comput., (25), 5-9 (2018).
9. L. Li, Q. M. Zhang, J. H. Zhao, & Y. W. Nie. Short-term traffic flow prediction method of different periods based on improved CNN-LSTM. J. Appl. Sci., **39**(2), 185-198 (2021).
10. X. Chen, J. Wang, & K. Xie. TrafficStream: A streaming traffic flow forecasting framework based on graph neural networks and continual learning. int. Joint Conf. Artif. Intell., (2021).T. Haarnoja, H. Tang, P. Abbeel, & S. Levine. Reinforcement Learning with Deep Energy-based Policies. In Proc. 35th Int. Conf. Mach. Learn., 1352-1361 (2018).
11. Y. Li, J. Hao, & D. Zha. Graph-based reinforcement learning for traffic signal control. Proc. AAAI Conf. Artif. Intell., **33**(01), 4167-4174 (2019).
12. F. Zheng, H. X. Liu, & F. Zheng. A reinforcement learning approach for adjusting dynamic traffic control parameters to manage traffic incidents. Transp. Res. C Emerg. Technol., **86**, 598-615 (2018).
13. O. Loyola-González. Black-box vs. White-box: Understanding their Advantages and Weaknesses from a Practical Point of View. IEEE Access, **7**, 154096-154113 (2019).
14. S. Chen, Z. He, & L. Sun. A DL Approach to the prediction of traffic flow with real-time data and its implementation in microservices. IEEE Trans. Intell. Transport. Syst., **22**(9), 5657-5668 (2021).