Research on the Application of Reinforcement Learning in Traffic Flow Prediction

Yiquan Hu

College of Intelligent and Computing, Tianjin University, Tianjin, 300354, China

Abstract. As global urbanization accelerates, urban traffic issues are becoming increasingly severe. Traffic flow prediction (TFP), as a key technology in intelligent transportation systems, aims to provide decision support and optimization plans by analyzing vehicle flow, speed, and density in road networks. However, traditional statistical models and prediction methods based on historical data exhibit many limitations when dealing with complex, dynamic, and nonlinear traffic flow data. The purpose of this paper is to discuss how Reinforcement Learning (RL) can be applied to TFP. RL optimizes strategies through interactions between agents and the environment to maximize cumulative rewards. High-dimensional state spaces and nonlinear problems can be handled with strong adaptability and strong adaptability. The paper provides a detailed review of the latest developments in Deep RL in the field of TFP, including the application of Q-learning and its variants in traffic signal control. Additionally, the article discusses the application of RL-based Long Short-Term Memory Networks, Graph Convolutional Networks (GCN), and Dynamic GCN in TFP. Although RL has achieved significant results in the field of TFP, its application still faces challenges such as data complexity, dynamics, and high computational resource consumption. The paper suggests that future research directions should include expanding abnormal data, improving model efficiency and scalability, and extending application scenarios to further advance intelligent transportation systems.

1 Introduction

As of 2019, 55% of the global population lived in urban areas, and this number is expected to be 13% by 2050 [1]. Despite significant developments in the construction of transportation facilities, the growth of the urban population has exacerbated traffic problems. The increasing demand for transportation by urban residents has led to decreasing traffic efficiency and rising accident rates, directly affecting the lives of citizens [2]. Traffic flow prediction (TFP) is a key technology in intelligent transportation systems. It analyzes and forecasts vehicle flow, speed, and density in road networks, providing decision support and optimization plans to enhance traffic efficiency and safety. Statistical models and historical data-based

Corresponding author: carrothuyq@tju.edu.cn

predictions, including time series analysis and regression models, are mainly used in traditional TFP methods. However, these methods face numerous limitations when handling complex, dynamic, and nonlinear traffic flow data, making them inadequate for dealing with emergencies and anomalies. As a result, exploring the use of more advanced technologies to enhance the accuracy and robustness of TFP has become a research focal point.

In recent years, with the swift advancement of artificial intelligence technologies, Reinforcement Learning (RL), as a dynamic decision-making approach, has demonstrated significant potential in the area of TFP. Unlike supervised learning, RL can autonomously learn strategies without requiring a significant amount of labeled data. Through the agent's engagement with the environment, RL continuously experiments and adjusts its strategies to maximize cumulative rewards. Compared to traditional methods, RL has strong adaptability and can handle high-dimensional state spaces and nonlinear problems. RL acquires experience through interaction with the environment and learns based on reward feedback. This means that RL agents can continuously improve their strategies over time, adapting to different environmental changes. This adaptability gives RL an advantage over traditional static methods in complex and dynamic environments [3].

Deep RL has been widely applied in TFP, with numerous RL algorithms such as Deep Q-Learning (DQN), Double Q-Learning, and Double DQN achieving remarkable results in TFP and traffic signal control [4-6]. TFP faces challenges such as dynamism, uncertainty, and complexity. RL addresses these by modeling problems as Markov decision processes, using Q-learning algorithms and deep RL algorithms to optimize strategies for traffic signal control, lane usage, and accident management, thereby significantly improving overall traffic efficiency. Multi-agent RL further enhances the collaborative prediction and optimization capabilities of large-scale traffic systems.

This paper reviews the latest research progress of deep RL in the field of TFP. It discusses the main challenges faced by the current application of RL in TFP and proposes future research directions to address these challenges. This paper aims to further advance the development of this field.

2 Research status analysis

2.1 Deep RL

Mnih first combined deep neural networks with Q-learning in 2013. The author employed raw pixels as input and a value function to estimate expected returns during the training of a convolutional neural network (CNN), which is a variant of Q-learning. In the Arcade Learning Environment, this model was evaluated in seven Atari 2600 games without adjusting the architecture or learning algorithm. The model outperformed all previous methods in six games and surpassed human experts in three games [7]. In 2015, Mnih expanded his work and first proposed the DQN algorithm. DRL achieved or exceeded human expert levels in multiple Atari 2600 games. DQN combined the Q-learning algorithm and deep neural networks to address control problems in high-dimensional state spaces. The introduction of DQN bridged the gap between high-dimensional input and action, creating the first artificial intelligence capable of learning and excelling in various challenging tasks [8]. Van proposed the Double DQN based on the Double Q-learning (DDQN) algorithm proposed in 2010 [9], addressing the overestimation problem of DQN.

This method altered the target value calculation by using the evaluation network to select the action and the target network for assessing the value. In experiments, the Double DQN algorithm with experience replay effectively mitigated the overestimation problem of the

DQN algorithm and improved prediction accuracy [10]. Wei proposed a DDQN-based intelligent system for traffic signal control, IntelliLight. Through multi-agent collaboration, it achieved signal optimization and TFP in complex urban traffic systems. Extensive use of synthetic and real-world data demonstrated that this system outperformed the current state-of-the-art signal optimization and TFP schemes [11].

2.2 RL Based Long Short-Term Memory Networks (RL-LSTM)

In the field of TFP, due to the ability of deep learning (DL) methods such as Recurrent Neural Networks (RNN) and their variant Long Short-Term Memory Networks (LSTM) to learn spatial-temporal dependencies, these methods have become excellent choices. However, they still have the drawbacks of having a wide variety of hyperparameters and being complex and difficult to design. Mostaf proposed an automated framework for TFP using LSTM based on RL design. High-performance LSTM predictors are generated by employing RL and Transfer Learning (TL) and this framework introduces a new DL algorithm called HERITOR (Higher-Order Traffic Convolution RL-LSTM) for TFP. To forecast future trajectories of moving objects lacking human intervention, a controller using Q-learning is developed to automatically design a high-performance LSTM architecture. The structure and parameters of the LSTM are optimized using the RL algorithm, and transfer learning is used to improve training efficiency, allowing the model to adaptively learn the optimal strategy. The HERITOR algorithm can extract features from the spatial structure of the traffic network through convolution operations, while utilizing LSTM to process time series data, better capturing the spatial-temporal characteristics of traffic flow. These captured spatial-temporal features are input into the RL-LSTM, achieving a more precise TFP in the spatial-temporal dimension. This method was tested on two real large-scale datasets, and the experimental results showed an improvement of 15% to 25% compared to the state-of-the-art technologies. Additionally, by transferring knowledge, the optimal LSTM architecture search process can be accelerated by up to 70% [12].

2.3 Dynamic Graph Convolutional Networks (DGCN) and RL based long-term TFP

To extract spatial-temporal features from traffic networks, many studies on TFP using DL technologies have combined CNN or Graph Convolutional Networks (GCN) [13-18]. Through the use of graphs, these methods effectively capture the topological structure and spatial correlations between road connections in the traffic network. However, real traffic networks are dynamic and dynamic graphs can better reflect the spatial-temporal characteristics of traffic compared to static graphs. In practical applications, the data is limited and timely, making it challenging to frequently gather statistics and generate graph structures. To address the issue of declining prediction accuracy in traditional TFP methods for long-term forecasting due to sparse and incomplete data, Peng proposed a long-term TFP method based on DGCN and RL. The system is represented by a dynamic traffic flow probability graph, with the model conducting convolution operations on the dynamic graph to capture spatial features and integrate LSTM units to learn temporal features. When the dynamic graph is insufficient due to sparse data, it is generated using a RL-based graph convolution strategy network. This model was experimented on the urban bicycle dataset in New York City and achieved high accuracy, enabling stable and effective long-term TFP, and reducing the impact of data deficiencies on prediction results [19].

2.4 GCN and RL

Currently, many studies in the field of TFP employ strategies using LSTM or a combination of LSTM and CNN [19-22]. However, these methods suffer from drawbacks such as high demand for training data, significant consumption of computational resources, or low model interpretability. Addressing these issues, Hang proposed a TFP method that combines GCN, LSTM, and RL to better reflect the dynamics of urban traffic flow and further optimize traffic control strategies. The model uses an RL-based GCN(RL-GCN) to process the urban traffic network's nodes and edges. First, the urban traffic network is modeled by GCN, representing it as a graph structure to derive the properties of each node and edge. After that, the LSTM network models the city's traffic flow and forecasts the future traffic flow. Finally, on this basis, a reward function and a policy network are designed, using RL to formulate optimal traffic control strategies. Experimental results on multiple real datasets indicate that, compared to previous studies, this method has higher model training efficiency and prediction accuracy, effectively predicting urban traffic flow and optimizing urban traffic flow efficiency [23].

3 Challenges

Despite the significant advancements RL has made in the field of TFP, its application still faces many challenges:

Insufficient Predictive Ability in Abnormal Situations: When making predictions in real traffic environments, the model needs to handle the interference of unexpected situations. Real traffic environments encounter abnormal situations such as traffic accidents and weather changes. However, such information is typically sparse and incomplete, leading to substantial room for improvement in the prediction accuracy of existing models under these conditions [24].

Complexity of Traffic Flow Data: Traffic flow data is complex because the state space may have high dimensionality. The problem of dimensionality disaster is faced by models when dealing with large-scale traffic networks, leading to low efficiency in training and inference [25].

High Demand for Computational Resources: Deep RL methods require repeated experimentation, consuming a large amount of computational resources and time. This leads to extremely high learning costs. For large-scale traffic networks, resource consumption increases significantly, severely impacting their application in real production environments [23].

4 Directions for further research

RL has considerable prospects for TFP. The development of intelligent transportation systems will be boosted by research on RL in this field. In response to the current challenges, the following points may become future research directions:

Generating Simulated Data Based on Generative Adversarial Networks (GAN): Using GAN to generate high-quality abnormal traffic flow data can enhance the robustness and accuracy of TFP models. By collecting and preprocessing traffic data, using GAN models for adversarial training, generating simulated abnormal data, and evaluating and optimizing, actual traffic anomalies can be effectively simulated. These generated data can be used for model training and testing, helping TFP systems identify and respond to potential anomalies, thereby improving the accuracy of TFP [26].

Improving Model Efficiency and Scalability: Adopting various mechanisms to improve the learning efficiency of the model. For example, conducting knowledge exchange between multiple intersections, as the operating environment of one intersection will affect the surroundings. Considering the traffic conditions of nearby intersections during the exchange and improving the global reward reflecting the state of the entire traffic network may help improve model efficiency [25].

Expanding Application Scenarios: Applying RL to more practical traffic management scenarios, such as route planning and vehicle-to-everything (V2X), to achieve a more comprehensive intelligent transportation management system.

5 Conclusion

The focus of this article is on the most recent research progress, challenges, and future development directions for RL in TFP. As a dynamic decision-making method, RL optimizes prediction effectiveness through continuous trial and strategy adjustment, demonstrating great potential in handling traffic flow dynamics and uncertainties. Methods that combine RL with DQN and its variants such as Double DQN, RNN, LSTM-based methods, DGCN, and GCN have significantly improved prediction accuracy and timeliness. However, current research still faces challenges such as the complexity and dynamics of traffic data, scarcity of abnormal data, and high computational resource requirements.

Although there are challenges, the scope of RL's applications in TFP is still very broad. Considering the current difficulties and challenges, this article suggests potential research directions for future RL applications in TFP. Continuous research efforts will provide crucial support and impetus for the progress of advanced and efficient traffic management solutions.

References

1. U. N. Desa. World Population Prospects 2019: Highlights. New York (US): United Nations Department for Economic and Social Affairs, **11**(1): 125 (2019).
2. B. Medina-Salgado, E. Sánchez-DelaCruz, P. Pozos-Parra, & J. E. Sierra. Urban traffic flow prediction techniques: A review. Sustain. Comput. Inform. Syst., **35**, 100739 (2022).
3. R. S. Sutton & A. G. Barto. Reinforcement Learning: An Introduction. MIT Press (2018).
4. H. Wang, Y. Yuan, X. T. Yang, T. Zhao, & Y. Liu. Deep Q learning-based traffic signal control algorithms: Model development and evaluation with field data. J. Intell. Transp. Syst., **27**(3), 314-334 (2023).
5. A. Agafonov & V. Myasnikov. Traffic signal control: A double q-learning approach. In 2021 16th Conference on Computer Science and Intelligence Systems (FedCSIS) (pp. 365-369) (2021).
6. N. T. T. Le, D. Tran-Anh, Q. B. Bui, & L. M. Pham. Double Deep Q-Network algorithm for solving traffic congestion on one-way highways. J. Water Resour. Environ. Eng., **87**(12), 39 (2023).
7. V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, & M. Riedmiller. Playing Atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602 (2013).

8. V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, ... & D. Hassabis. Human-level control through deep reinforcement learning. Nature, **518**(7540), 529-533 (2015).
9. H. Hasselt. Double Q-learning. Adv. Neural Inf. Process. Syst., **23** (2010).
10. H. Van Hasselt, A. Guez, & D. Silver. Deep reinforcement learning with double Q-learning. Proc. AAAI Conf. Artif. Intell., **30**(1) (2016).
11. H. Wei, G. Zheng, H. Yao, & Z. Li. Intellilight: A reinforcement learning approach for intelligent traffic light control. In Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (pp. 2496-2505) (2018).
12. M. Karimzadeh, R. Aebi, A. M. de Souza, Z. Zhao, T. Braun, S. Sargento, & L. Villas. Reinforcement learning-designed LSTM for trajectory and traffic flow prediction. In 2021 IEEE Wireless Commun. Networking Conf. (WCNC) (pp. 1-6) (2021).
13. Z. He, C. Y. Chow, & J. D. Zhang. STCNN: A spatio-temporal convolutional neural network for long-term traffic prediction. In 2019 20th IEEE Int. Conf. Mobile Data Manage. (MDM) (pp. 226-233) (2019).
14. Y. Liu, H. Zheng, X. Feng, & Z. Chen. Short-term traffic flow prediction with Conv-LSTM. In 2017 9th Int. Conf. Wireless Commun. Signal Process. (WCSP) (pp. 1-6) (2017).
15. B. Du, H. Peng, S. Wang, M. Z. A. Bhuiyan, L. Wang, Q. Gong, ... & J. Li. Deep irregular convolutional residual LSTM for urban traffic passenger flows prediction. IEEE Trans. Intell. Transp. Syst., **21**(3), 972-985 (2019).
16. Z. Cui, K. Henrickson, R. Ke, & Y. Wang. Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. IEEE Trans. Intell. Transp. Syst., **21**(11), 4883-4894 (2019).
17. L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, ... & H. Li. T-GCN: A temporal graph convolutional network for traffic prediction. IEEE Trans. Intell. Transp. Syst., **21**(9), 3848-3858 (2019).
18. B. Yu, H. Yin, & Z. Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875 (2017).
19. H. Peng, B. Du, M. Liu, M. Liu, S. Ji, S. Wang, ... & L. He. Dynamic graph convolutional network for long-term traffic flow prediction with reinforcement learning. Inf. Sci., **578**, 401-416 (2021).
20. T. Zhang & G. Guo. Graph attention LSTM: A spatiotemporal approach for traffic flow forecasting. IEEE Intell. Transp. Syst. Mag., **14**(2), 190-196 (2020).
21. L. Cai, M. Lei, S. Zhang, Y. Yu, T. Zhou, & J. Qin. A noise-immune LSTM network for short-term traffic flow forecasting. Chaos: An Interdiscip. J. Nonlinear Sci., **30**(2) (2020).
22. H. Zheng, F. Lin, X. Feng, & Y. Chen. A hybrid deep learning model with attention-based Conv-LSTM networks for short-term traffic flow prediction. IEEE Trans. Intell. Transp. Syst., **22**(11), 6910-6920 (2020).
23. H. Xing, A. Chen, & X. Zhang. RL-GCN: Traffic flow prediction based on graph convolution and reinforcement learning for smart cities. Displays, **80**, 102513 (2023).
24. P. Ø. Kanestrøm. Traffic flow forecasting with deep learning. Master's thesis, NTNU (2017).

25. F. Rasheed, K. L. A. Yau, R. M. Noor, C. Wu, & Y. C. Low. Deep reinforcement learning for traffic signal control: A review. IEEE Access, **8**, 208016-208044 (2020).
26. C. Wu, L. Chen, G. Wang, S. Chai, H. Jiang, J. Peng, & Z. Hong. Spatiotemporal scenario generation of traffic flow based on LSTM-GAN. IEEE Access, **8**, 186191-186198 (2020).