Frontiers in Traffic and Transportation Engineering (2025)
Clausius Scientific Press, Canada
DOI: 10.23977/ftte.2025.050105
ISSN 2616-2334 Vol. 5 Num. 1

Traffic Congestion Pricing Mechanism Based on Deep Reinforcement Learning

Ying Wang[1,a*], Hongchun Zhang[2,b]

[1]Tandon School of Engineering, New York University, Brooklyn, 11201, New York, United States
[2]School of Transportation, Beijing Jiaotong University, Haidian District, 100080, Beijing, China
[a]yw8220@nyu.edu, [b]19251218@bjtu.edu.cn
**Corresponding author*

Keywords: Deep Q network, congestion dynamic pricing, reinforcement learning

Abstract: With the breakthrough progress of deep reinforcement learning technology, intelligent traffic governance has become the core research direction of urban management. To solve the problem that traditional congestion pricing mechanism is difficult to respond dynamically to the changes of road network state, this paper proposes a dynamic pricing model based on deep Q network. Through reinforcement learning framework, the model takes real-time traffic state as input, combines experience replay and objective network mechanism to generate pricing policy suitable for spatiotemporal characteristics. Experimental results show that DQN performs best in cumulative reward and discount cumulative reward, and converges faster by about 20% compared with fixed rate model and variable pricing model. The experimental model achieves a 12% increase in emission reduction rate while ensuring fiscal revenue through a collaborative strategy of peak online ride-hailing premium and bus discount. This research provides a real-time responsive decision-making framework for urban smart transportation governance, helping to break through the efficiency bottleneck of the traditional static pricing model.

1. Introduction

In recent years, breakthrough progress in deep reinforcement learning technology has provided a new technical path for urban traffic congestion management. Traditional congestion pricing mechanisms are mainly based on fixed rates or simple floating rules, which are difficult to dynamically adapt to the real-time changes of road network traffic state and the complex differences of traveler behavior. With the breakthrough of reinforcement learning algorithm, deep Q network (DQN) shows significant advantages in the field of decision optimization. It realizes policy optimization through interaction with environment, which provides theoretical basis for constructing dynamic response pricing policy model [1]. The current traffic pricing system generally adopts static or pre-programmed dynamic policies. Although it can initially realize the basic rate setting, it is difficult to adjust the pricing policy according to the real-time congestion distribution and travel demand of the road network. DQN algorithm characterizes high-dimensional traffic state through deep neural network, and combines experience replay and target network mechanism, which can effectively solve the stability problem in the process of policy training, and provides technical support for continuous optimization of dynamic pricing policy.

This paper proposes a congestion pricing mechanism based on DQN, aiming at realizing dynamic generation and continuous evolution of pricing strategy through reinforcement learning framework, so as to help urban traffic managers break through the efficiency bottleneck of traditional static pricing model.

2. DQN Algorithm

Deep Q network algorithm is an important breakthrough in the field of reinforcement learning. It can effectively overcome the dimensional disaster problem of traditional Q-learning in high-dimensional state space by approximating Q function through deep neural network. As shown in Figure 1, DQN takes the environment state as input, and the agent chooses to perform actions and interact with the environment according to the current policy. The environment then feeds back new status and instant reward signals. The key step is to calculate the difference between the predicted Q value and the actual reward signal to form a loss function, and update the neural network parameters by back propagation to achieve iterative optimization of the policy [2].

In the implementation of DQN algorithm, accurate representation of state space constitutes the basis of decision. The agent evaluates optional actions in the action space based on the current state, triggers environmental state transitions by executing actions, and obtains reward feedback. The design of reward function directly affects the direction of strategy optimization, and short-term goal achievement and long-term system stability should be considered comprehensively. In order to improve the quality of decision making, the algorithm usually adopts ε-greedy strategy to balance the relationship between exploring new actions and utilizing existing knowledge.

Q-value function approximation is the core technical contribution of DQN. Q value represents the expected value of long-term cumulative reward obtained by executing an action in a specific state. Traditional Q-learning relies on tables to store Q values of discrete states, which is difficult to deal with high-dimensional continuous state space. DQN innovatively adopts deep neural network as function approximator, and uses its strong nonlinear fitting ability to establish the mapping relationship from state to action to Q value [3]. The loss function is usually constructed by the mean square error between the predicted Q value and the target Q value, and the network parameters are gradually optimized by gradient descent algorithm. This design makes DQN exhibit excellent adaptability in decision-making tasks in discrete action space, laying a technical foundation for its application in robot control, resource scheduling and other fields.

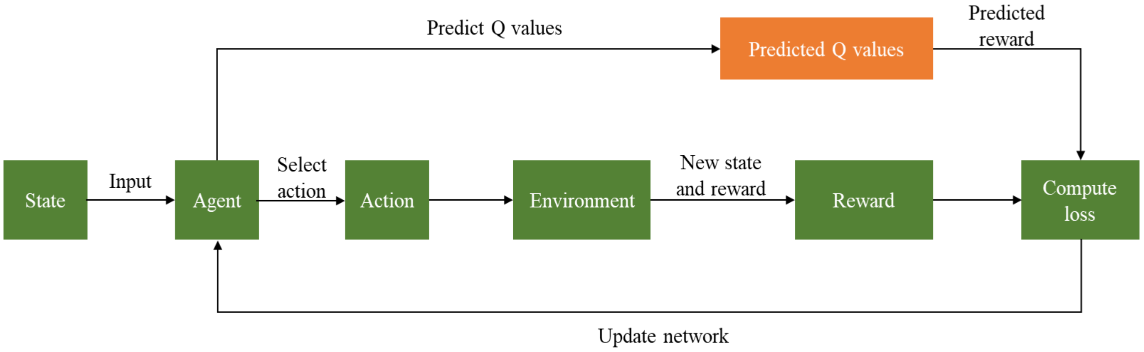

Figure 1 Schematic diagram of DQN

3. Algorithm of Traffic Congestion Pricing Mechanism

3.1. Model Frame Design

The congestion pricing framework consists of four functional modules: road network state perception, pricing agent decision, traffic system interaction and policy optimization, forming a "state-action-feedback-update" closed-loop decision system. The model takes real-time traffic state as input starting point, covers multi-dimensional features such as regional traffic density, road travel speed, traffic demand elasticity, etc., and is transformed into high-dimensional vector representation by spatial-temporal feature encoder to provide decision-making basis for agents. The agent module constructs a policy network based on a deep reinforcement learning algorithm, and outputs the probability distribution of the zonal dynamic pricing policy after receiving the state vector [4]. The specific actions include rate grade adjustment, preferential period setting, congestion area division and other regulation means.

After the pricing strategy is executed, it enters the traffic environment module, which simulates the driver's response process to the price signal according to the travel behavior theory. The new state generation mechanism integrates traffic flow dynamics and trip selection models to dynamically calculate path transition probability and demand elasticity under different tariffs, ensuring scientific and interpretable state evolution. At the same time, the environment module calculates the instant reward value according to the preset social benefit function, and the reward design needs to balance short-term congestion alleviation and long-term system benefit, and realize policy-oriented precise regulation through multi-objective weighting.

The strategy value evaluation module models the long-term social benefit of state-strategy pairs through deep neural networks, and outputs a predictive value function as the basis for strategy evaluation. This module adopts dual network architecture, online policy network is responsible for generating real-time value estimation, target value network provides stable benchmark value for loss calculation, and their parameters are updated asynchronously to mitigate value fluctuation. The loss function is constructed by comparing the mean square error of the discounted value of the current benefit superimposed on the future benefit. The gradient signal updates the network parameters through backpropagation and drives the continuous optimization of the pricing strategy.

3.2. Model Training Methods

The core goal of model training is to make agents dynamically generate optimal pricing policies that adapt to real-time traffic conditions by iteratively optimizing the parameters of policy networks. The training process is divided into three stages: network initialization, experience collection, strategy updating and parameter synchronization.

The experience playback pool capacity is set to the order of 1e5 to 1e7 depending on the road network size, ensuring storage of diverse traffic state data and supporting efficient random sampling. During the training process, the agent selects pricing actions based on the ε-greedy policy. After each action is executed, the system stores the quadruple (road network state, pricing action, economic environment feedback, new state) into the experience replay pool, and initiates the network update when the data volume reaches the preset batch size. In the sampling process, priority experience replay technique is used to dynamically allocate sample weights according to time series difference errors, and high value decision experience is trained first to accelerate convergence.

The Huber loss function is used to balance the stability and sensitivity of training. The target Q value is calculated by the target network: for each sample, the target Q value is defined as the discounted value of the immediate economic benefit and the long-term road network benefit, i.e., the benefit function needs to integrate the fare revenue, congestion mitigation degree and external costs (such as carbon emissions). The prediction network minimizes the difference between prediction Q and target Q by gradient back propagation, and the optimizer adopts RMSProp algorithm to adjust learning rate adaptively to avoid local optimization trap. Target network parameters are predicted synchronously at fixed training steps to ensure the stability of policy evaluation.

This training framework maps dynamic traffic state into continuous decision space, solves the problem of policy oscillation and slow convergence of traditional pricing model in time-varying road network through dual network architecture and priority sampling mechanism, and provides an extensible technical path for real-time congestion pricing.

3.3. Model Performance Evaluation

In order to evaluate the performance of congestion pricing model based on deep reinforcement learning, cumulative reward, average reward and discounted cumulative reward are selected as evaluation system to measure the economic benefit, stability and long-term social value of pricing strategy.

Cumulative reward reflects the total revenue of agent in the whole training cycle, which is the core index to measure the comprehensive effect of pricing strategy. In the traffic management scenario, the index integrates multiple objectives such as ticket revenue, road network traffic efficiency and emission reduction effect, and quantifies the global optimization degree of the strategy through the weighted design of reward function [5]. A higher cumulative reward indicates that the model is able to dynamically balance short-term economic gains (e.g., peak-hour premiums) with long-term sustainability goals (e.g., carbon emissions control).

Average reward evaluates the stability of a pricing strategy by calculating the average reward for a single training run. Traffic system has significant spatiotemporal fluctuation and sudden disturbance (such as accident and weather). Stable average reward value indicates that the model can adapt to the change of road network state in different periods and avoid strategy shock caused by sudden demand change or external shock. This metric particularly validates the robustness of the model in unsteady scenarios such as holiday scheduling and extreme weather response.

The discounted cumulative reward emphasizes the contribution of recent pricing actions and assesses the long-term planning ability of the model for social benefits. The implementation of transportation policy has continuous transmission effect, which can quantify the effectiveness of the model in demand elasticity management and low-carbon travel guidance. Higher discount incentives indicate that the model can reasonably plan the rate adjustment rhythm, avoid public resistance or market imbalance caused by short-term aggressive price adjustment, and ensure policy sustainability.

This evaluation system relates pricing strategy and system response through quantitative indicators, breaks through the limitation of traditional focus only on revenue or single congestion coefficient, and provides multidimensional criteria for closed-loop optimization of dynamic pricing model.

4. Empirical Analysis

4.1. Experimental Design

To verify the effectiveness of deep reinforcement learning algorithm in traffic congestion pricing, this study selects fixed-rate policy and demand-based floating pricing model as comparison benchmarks, focusing on the difference between real-time response ability and long-term policy sustainability.

The experimental parameters are classified into three categories: algorithm parameters, environment parameters and traveler parameters to adapt to the dynamic complexity of urban traffic system. DQN algorithm adopts three-layer fully connected neural network, the input layer dimension is consistent with the road network state characteristics (including regional traffic density, average speed, etc.), the number of hidden layer neurons is adaptively adjusted according to the road network scale (64-256 units), and the output layer is mapped to the rate adjustment action space. The flat-rate model sets flat prices based on historical averages; the floating-rate model adjusts rates linearly based on real-time traffic indices (with a float factor of 0.1-0.3).

The experiment simulates the data of 8000 vehicles and compares the performance of the algorithm in cumulative road network revenue, average traffic efficiency and policy convergence efficiency. In the training phase, a gradual scenario loading strategy is adopted, which simulates flat peak road conditions on weekdays at the initial stage, and gradually introduces peak congestion and extreme weather events to test the robustness of the algorithm. Fixed random seed was used to eliminate random interference in evaluation phase, and Monte Carlo sampling was used to verify the stability of results.

4.2. Experimental Results and Analysis

Table 1 shows the performance comparison results of different pricing algorithms on the test set. Deep reinforcement learning algorithms significantly outperform benchmark models in cumulative rewards and discounted cumulative rewards. When γ=0.95, the average reward fluctuates at the initial stage due to the decay mechanism of exploration rate, and gradually stabilizes with the training progress. The fixed-rate model performs better in stable road network environment, while the floating-rate model increases the reward by 12.6% in peak scenario due to its real-time response ability. The cumulative benefit advantage of DQN stems from the effective integration of experience replay and target network for long-term social benefits. It should be noted that the fluctuation range of DQN average reward is significantly larger than that of fixed rate model, which is mainly due to the active attempt of differential pricing strategy in early exploration stage.

Table 1 Experimental Result

Model	Cumulative reward	Average reward	Discount cumulative rewards
DQN Dynamic Pricing	1,528.7	15.3	2,043.2
Fixed rate model	1,212.4	12.1	1,568.9
Floating pricing model	1,376.5	13.7	1,802.6

Convergence analysis shows that DQN converges within 700 steps. The reward fluctuation in the exploration phase reflects the active exploration characteristic of ε-greedy strategy, while the fast stability in the later phase confirms the adaptability of the algorithm to dynamic traffic environment.

The multi-objective optimization results further verify the superiority of DQN. Through the coordinated regulation of peak online car-booking premium and bus discount, the emission reduction rate can be increased by 12% while the revenue loss is controlled within 5% under the sudden passenger flow scenario. This exact balance results from the weighted integration of economic benefits and carbon emissions indicators in the reward function, and from the accurate discounting of long-term policy benefits by the target network.

5. Conclusion

A dynamic pricing model based on reinforcement learning is constructed by introducing deep Q network into traffic congestion management field. The model realizes multi-objective optimization of pricing strategy by real-time perception of road network state, dynamic generation of pricing action, and simulation of policy transmission effect by combining traffic flow dynamics and trip choice theory. Experimental results show that DQN is significantly superior to fixed rate and variable pricing models in cumulative reward and long-term social benefit, and its convergence speed is improved. The experimental model improves emission reduction efficiency while ensuring fiscal revenue through the collaborative strategy of online ride-hailing premium and bus discount during peak hours. The research not only expands the application boundary of deep reinforcement learning in public policy formulation, but also provides a scalable technical framework for urban intelligent transportation governance.

References

[1] Jakob, M., Menendez, M. Parking pricing vs. congestion pricing: a macroscopic analysis of their impact on traffic[J]. Transportmetrica A: Transport Science, Vol. 17 (2021) No. 4: 462-491.

[2] González-Aliste, P., Derpich, I., López, M. Reducing urban traffic congestion via charging price[J]. Sustainability, Vol. 15 (2023) No. 3: 2086.

[3] Li, W., Kockelman, K. M., Huang, Y. Traffic and welfare impacts of credit-based congestion pricing applications: An Austin case study[J]. Transportation Research Record, Vol. 2675 (2021) No. 1: 10-24.

[4] Hall, J. D. Can tolling help everyone? Estimating the aggregate and distributional consequences of congestion pricing[J]. Journal of the European Economic Association, Vol. 19 (2021) No. 1: 441-474.

[5] Cohen-Blankshtain, G., Bar-Gera, H., Shiftan, Y. Congestion pricing and positive incentives: conceptual analysis and empirical findings from Israel[J]. Transportation, Vol. 50 (2023) No. 2: 607-633.