European Journal of AI,
Computing & Informatics

Vol. 1 No.1 2025

Article **Open Access**

Innovative Application of Reinforcement Learning in User Growth and Behavior Prediction

Zhuoer Ma [1,*]

[1] Acorns, Analytics, Irvine, California, 92617, United States
* Correspondence: Zhuoer Ma, Acorns, Analytics, Irvine, California, 92617, United States

Received: 09 March 2025
Revised: 15 March 2025
Accepted: 27 March 2025
Published: 02 April 2025

Abstract: With the rapid development of Internet technology, more and more industries realize the importance of user scale and user prediction. Although the traditional user prediction methods have achieved certain results in some specific scenarios, they generally have the disadvantages of inaccurate prediction and inadaptability to the changes of scenarios. In recent years, due to the characteristics of autonomous learning and strong adaptability, machine learning technology based on reinforcement learning has broad application prospects in personalized recommendation system, multi task optimization, user behavior prediction and so on. The focus of this paper is on the means and methods to help expand the scale of users and improve the ability of behavior prediction through reinforcement learning. This includes the establishment of personalized recommendation based on reinforcement learning; combining multi task learning with Multi-Agent Reinforcement Learning; a novel method combining deep reinforcement learning and behavior sequence prediction is studied. This paper analyzes the current situation of reinforcement learning in this field, and puts forward innovative strategies to further optimize the existing model, so as to better improve the real-time and adaptability. This paper provides a new idea for the application of reinforcement learning assisted behavior prediction, and also lays a theoretical foundation for future related work.

Keywords: reinforcement learning; personalized recommendation; multi task learning; user behavior prediction

1. Introduction

In the network era, making full use of the user growth strategies and behavior prediction models to improve the operating efficiency of websites or services has become the main focus of many companies and research. Traditional methods infer user behavior patterns by analyzing historical information, but they are often ineffective for massive information, complex user activities and continuous change process. Especially for personalized recommendation, estimation of user retention rate and prediction of behavior sequence, various models also have many difficulties, such as cold start problem, conflict between objectives, long-term dependence problem, etc. Reinforcement learning, as a self-regulated learning ability, continuously optimizes its decision-making process through the interaction between agents and the environment, and has been widely used in recommendation systems, advertising, game AI and other fields. In the process of user growth and behavior prediction, reinforcement learning can dynamically adjust individual prediction strategies through instant response to achieve more accurate prediction accuracy and more personalized recommendation. In addition, reinforcement learning can handle multiple optimization objectives at the same time through the multi task learning framework and multi-agent system, which solves the conflict between a series of objectives and improves the overall efficiency.

Although the application of reinforcement learning in this field is still in its infancy, there are still some problems to be solved, such as how to better reflect the changes of user behavior patterns, how to carry out rapid training of large quantities of data, and so on. Therefore, this paper conducts an in-depth analysis of the practical applications and innovative strategies of reinforcement learning in the increase of the number of users and behavior prediction, and introduce the application of new technologies related to this method in the fields of personalized recommendation, goal setting adjustment and behavior sequence prediction, and provide theoretical reference and practical guidance for future related research work.

2. Theoretical Framework of Reinforcement Learning

Reinforcement learning (RL), as a kind of machine learning, is a trial and error learning based on trying to make mistakes. It is mainly based on the interaction between the agent and the environment, and learns the optimal strategy. Its core idea is to enable agents to learn the best behavior strategy of accumulating rewards from the environment they are in, rather than being trained with supervised information. The structure of reinforcement learning is shown in Figure 1.

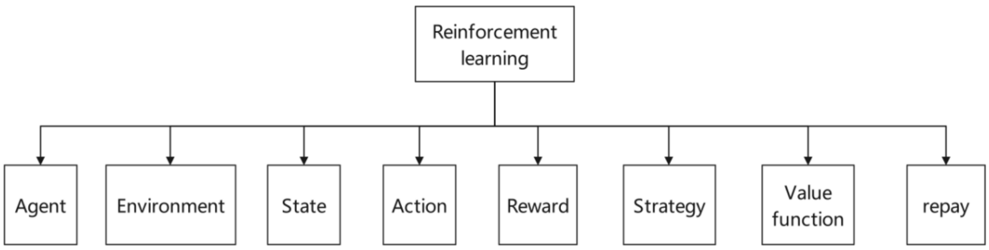

Figure 1. framework of reinforcement learning.

The goal of reinforcement learning is to optimize strategies in the interaction between agents and the environment to maximize the cumulative rewards. Among them, there are mainly the following reinforcement learning algorithms: value iteration method, according to the method of dynamic programming, through iterative update until approaching the optimal strategy; Q-learning, a model-free RL technique, guides the agent to take the optimal behavior by learning a state action value function $Q\ (s, a)$. The advantage of this algorithm is that it does not need a preset model for the environment, and it can obtain the best strategy from the environment; Deep q-network (dqn), the extension of Q-learning, uses deep learning neural network to construct the Q-value function. Dqn extends the application scenario of reinforcement learning to problems in high-dimensional state space, such as video games, robot control, etc. The policy gradient methods can directly modify the policy and bypass the process of improving the value function. It is suitable for larger continuous action spaces and is often used in the design of more difficult problems such as robot control and game AI [1].

3. Current Situation of Reinforcement Learning in User Growth and Behavior Prediction

3.1. Insufficient Accuracy and Cold Start of Personalized Recommendation

The purpose of personalized recommendation is to provide content or items suitable for the target user according to their behavior, preferences and interests. However, the current personalized recommendation methods have a series of problems in accuracy and coverage, especially in poor accuracy, cold start problems and so on.

In terms of the lack of accuracy, Due to the limitations of collaborative filtering and content-based recommendation methods, the traditional personalized recommendation algorithms lack the guarantee of accuracy, and when the user's information is less, it often leads to the reduction of the quality of recommendation results. For example, collaborative filtering finds similar users or similar products from the perspective of users' preferences and habits. It is difficult to identify the exact interests of low-frequency users, and there is a risk of recommendation errors. In addition, the complexity of users' preferences, coupled with the rigidity of the recommendation system, has seriously weakened the possibility of accurate matching.

Another factor that will affect the performance of the recommended system is the cold start problem. Cold Start refers to the situation that occurs when new users or new products are encountered in the recommendation system. Due to the lack of sufficient information about the behavior data of new users, it is difficult to analyze their preferences and needs, so it is difficult to get accurate recommendation results. When encountering new products, it is also difficult to make effective recommendations for relevant users due to the lack of relevant information about other products that can be referenced and compared; Although the hybrid recommendation strategy is adopted in the existing recommendation system, the positioning of new users and new product information still need to be accurately positioned in the cold start, which also brings great difficulties to the personalized recommendation system, and is one of the most important problems in the cold start [2].

3.2. Goal Conflict and Optimization Bottleneck in Multi Task Learning

For the tasks of user growth and behavior prediction, multi task learning (MTL) is widely used in the simultaneous output of multiple tasks, but the problems of goal conflict and optimization bottleneck often occur in the practical application of MTL.

Using the common training process for multi task learning to improve the common learning method of multiple tasks, we usually hope to use the effect of cooperation between different tasks to improve the overall performance of the model. However, in reality, there are often objective contradictions between different tasks. For example, in the case of a recommendation system, in order to improve the short-term hits and long-term user retention at the same time, there will be a conflict of goals. For the short-term recommendation system, it is more inclined to recommend more attractive web content to stimulate users' click motivation, but in the long run, it is also more likely that users will exit due to over recommendation of low-quality content.

In addition, the bottleneck phenomenon needs to be optimized in the context of multi task learning, because all tasks share a common model in multi-task learning, so the gradient and update between different tasks may interfere with each other. For example, when the amount of calculation required for a task is high, it will occupy a lot of system resources, resulting in the phenomenon that the learning process of other tasks will be affected. Especially in the case of large differences in the difficulty of different tasks, one task will occupy a large position in the whole training process, but other tasks will not be trained. This kind of bottleneck phenomenon will reduce the performance of the overall model and aggravate the instability in the training phase.

3.3. Dynamic Modeling and Long-Term Dependence in User Behavior Prediction

User behavior prediction is an important part of personalized recommendation and user growth. However, in practice, due to the challenges of dynamic modeling and long-term dependence, the prediction ability and adaptability of the current model are significantly affected.

For the problem of dynamic modeling, it means that user behavior will change with the change of time, external factors and many other variables, but the existing prediction models are not good at dealing with such dynamics in real time. Most conventional behavior prediction methods only use historical data and ignore the relationship between user behavior and time. Therefore, such a model cannot respond to changes in user preferences, especially when the form of user behavior suddenly changes, the accuracy of its user behavior prediction will also decline sharply.

In some cases, some behaviors of users may have a certain correlation with some behaviors of users a long time ago, which is called "long-term dependence problem". For example, the inactivity of user behavior at a certain stage may be due to the transfer of long-term interest, but this important information has not been modeled into the model. Time series prediction technology often encounters great difficulties in solving such problems, mainly because time series models are good at grasping the short-term correlation, but not good at maximizing the role of long-term historical information in predicting future behavior [3].

4. Innovation of Reinforcement Learning in User Growth and Behavior Prediction

4.1. Personalized Recommendation Algorithm Based on Reinforcement Learning

Personalized recommendation system has been widely used in e-commerce, audio-visual media, social networks and many other fields. The traditional recommendation algorithm based on the static modeling of the user's previous activity records is easy to cause the attenuation of accuracy, while reinforcement learning can update the recommendation strategy in real time according to the user's feedback, which can significantly improve the accuracy of recommendation. For example, a shopping website used Q-learning algorithm to optimize the commodity recommendation system. In this approach, the state is the user's historical activity data, the action is the product recommended to the user, and the reward is the user's click or consumption. The updated formula is:

$$Q(s\_t, a\_t) = Q(s\_t, a\_t) + \alpha[(R\_(t+1) + \gamma\ \max\!\!\!-a'\,Q\ \ (s\_(t+1), a') - Q(s\_t, a\_t)] \quad (1)$$

Here, s_t and a_t respectively represents the status and action at the current time, where $R\_(t+1)$ represents the reward after the action is executed, γ is the discount factor, which indicates the importance of future rewards, and α is the learning rate. It can be seen from this that in the process of intensive learning and training, we can observe the impact of recommended goods on user behavior, find and use such a balance point to modify the way of commodity recommendation, and constantly try to optimize commodity recommendation, thus gradually improving the click-through rate and purchase conversion rate of products [4]. The data comparison between traditional algorithm and reinforcement learning algorithm is shown in Table 1.

Table 1. Data comparison between traditional algorithm and reinforcement learning algorithm.

Cycle time	Traditional recommendation algorithm (click through rate)	Reinforcement learning recommendation algorithm (click through rate)	Traditional recommendation algorithm (purchase conversion rate)	Reinforcement learning recommendation algorithm (purchase conversion rate)
1 week	5%	7%	1.5%	2%
2 weeks	6.5%	8.5%	2%	3%
3 weeks	7%	9.5%	2.2%	4%
4 weeks	7.5%	10%	2.5%	5%

It can be seen from the data in the table above that the click through rate and purchase conversion rate of the reinforcement learning recommendation algorithm after a long learning process are gradually better than those of the traditional technology. More obviously, after the fourth cycle, the click through rate and purchase conversion rate are higher than 2.5 percentage points, indicating that the reinforcement learning model can be timely optimized based on real-time user behavior feedback to obtain more accurate personalized services. Compared with the fixed traditional recommendation method, reinforcement learning model can better adapt to the changes of users' personalized characteristics, so as to improve the long-term effect.

4.2. Multi-Task Learning and Multi-Agent Reinforcement Learning

Multi task learning (MTL) and Multi-Agent Reinforcement Learning (MAS) are used to process multiple objectives in the same model, so as to optimize the overall effect of user growth and behavior prediction. In practical application, these technologies solve the conflict problem between multiple objectives. For example, in a social platform, the multi task learning model is used to optimize the following three goals at the same time: (1) improve user retention rate, (2) increase user activity, and (3) improve user purchase conversion rate. The formula of the objective function is:

$$L = \omega\_1 \cdot L\_retention + \omega\_2 \cdot L\_engagement + \omega\_3 \cdot L\_conversion \quad (2)$$

Where,$L_{retention}$ is the retention rate,$L_{engagement}$ is the activity,$L_{conversion}$ is the purchase conversion rate, ω_1、ω_2 and ω_3 are the weight factor, which represent the contribution of retention rate, activity, and purchase conversion rate to the system optimization goal Therefore, the platform constructs a system based on Multi-Agent Reinforcement Learning, and each agent is responsible for a specific task. These agents share some network layers and work together to improve the efficiency of the system. At the same time, the reward and punishment mechanism of each agent is adjusted differently according to the type, importance and impact of different tasks, so as to achieve the best balance and coordination under multiple objectives [5]. The data comparison between traditional algorithm and reinforcement learning algorithm is shown in Table 2.

Table 2. Data comparison between traditional algorithm and reinforcement learning algorithm.

Cycle time	Traditional algorithm (retention rate)	Reinforcement learning multi task algorithm (retention rate)	Traditional algorithm (activity)	Reinforcement learning multi-tasking algorithm (activity)	Traditional algorithm (purchase conversion rate)	Reinforcement learning multi task algorithm (purchase conversion rate)
1 week	45%	50%	60%	65%	3%	4%
2 weeks	48%	54%	62%	68%	3.5%	5%
3 weeks	50%	58%	64%	70%	4%	6%
4 weeks	52%	62%	66%	72%	4.5%	7%

It can be found from the table that the multi task method of reinforcement learning has a significant impact on the user retention rate, activity participation and purchase conversion rate. Taking the fourth week as an example, the impact on the user retention rate, activity participation and purchase conversion rate is 10%, 6% and 2.5% respectively. Reinforcement learning can better balance different task objectives and use different agents to coordinate and adjust each other, and prevent the conflict between objectives and Optimization in traditional methods.

4.3. Combination of Deep Reinforcement Learning and Behavior Sequence Prediction

The prediction of user behavior sequence is to predict the future behavior through the historical behavior data of users, which is widely used in e-commerce, social networks and content recommendation. Traditional prediction methods usually use time series as prediction means, but they can not deal with the long-span correlation and changing user preferences. Deep reinforcement learning (DRL) is a method combining deep neural network and reinforcement learning model, which can better mine the long-term dependence in user behavior, so as to effectively break through the traditional constraints. This method identifies high-order characteristics through deep neural networks (such as convolutional neural networks, long-term and short-term memory networks, etc.), and then combines with the strategy adjustment of reinforcement learning, which can comprehensively consider the long-term dependence of user behavior and personalized needs in the process of dealing with dynamic behavior sequence [6]. The operation process of deep reinforcement learning in behavior sequence prediction is shown in Figure 2.

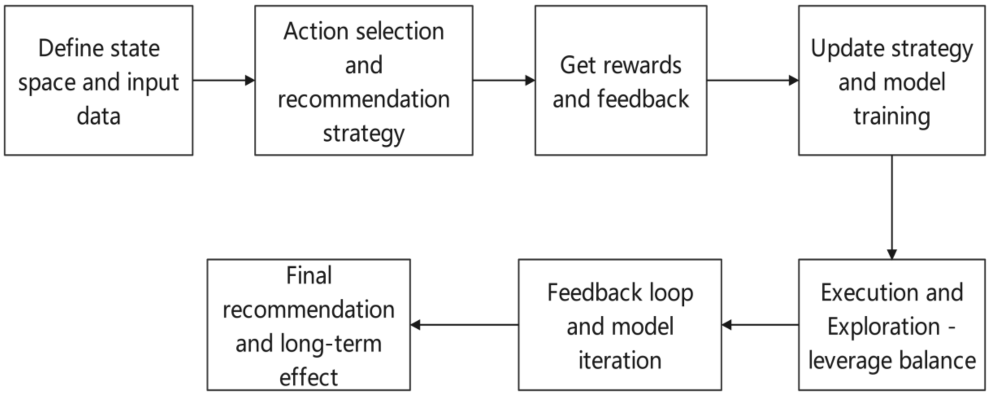

Figure 2. operation process of deep reinforcement learning in behavior sequence prediction.

Taking a video app user behavior prediction as an example, the user's viewing records, browsing data and retrieval data are all the user's behavior tracks. Traditional methods may use simple regression model (such as ARIMA model) or recommendation system based on collaborative filtering, but it is difficult to capture the complexity and dynamics of user behavior (for example, the user's preference for a certain film will gradually appear in a few weeks). Deep reinforcement learning can simulate this long-term impact relationship and accurately predict the potential interest of users in the future. Specifically, the video app uses dqn or strategy gradient method to improve its recommendation system. Taking the action track of a single user as the input, the neural network is used to capture the user's interest characteristics. Then, reinforcement learning changes the recommendation scheme by taking the optimal recommendation steps in different environments (i.e. action information at different stages), so as to achieve more personalized and accurate prediction.

5. Conclusion

With the continuous progress of network and artificial intelligence technology, reinforcement learning has shown great potential to bring user growth and predict user behavior for services. Reinforcement learning can significantly improve the accuracy and real-time of personalized recommendation through the reasonable optimization of recommendation scheme, the solution of long-term dependence problem and the treatment of high-level complexity of user behavior. This paper studies the application of reinforcement learning in new fields, such as personalized recommendation, multi task learning, behavior sequence prediction and others, and also discusses the advantages compared with traditional methods. In the future, with the progress of algorithms and the application of big data technology, reinforcement learning may have a greater impact on population growth and behavior prediction, bringing accurate and efficient solutions to all walks of life.

References

1. B. R. Chang, H. F. Tsai, and G. R. Chen, "Self-Adaptive Server Anomaly Detection Using Ensemble Meta-Reinforcement Learning," *Electronics*, vol. 13, no. 12, p. 2348, 2024, doi: 10.3390/electronics13122348.
2. H. Zhang, P. Zhao, and H. Z. Cui, "Click is not equal to purchase: multi-task reinforcement learning for multi-behavior recommendation," *World Wide Web*, vol. 26, no. 6, pp. 4153-4172, 2023, doi: 10.1007/s11280-023-01215-6.
3. S. Wan, S. Yang, and Z. Fu, "Focus on user micro multi-behavioral states: Time-sensitive user behavior conversion prediction and multi-view reinforcement learning based recommendation approach," *Inf. Process. Manag.*, vol. 62, no. 2, p. 103967, 2025, doi: 10.1016/j.ipm.2024.103967.
4. G. Yang, K. Li, Y. Zhang, and Q. Feng, "End-to-End Modeling and Long Short-Term Memory Application in Time Series Modeling: Modeling and Prediction of Electronic Commerce User Behavior," *J. Organ. End User Comput.*, vol. 36, no. 1, pp. 1-27, 2024, doi: 10.4018/JOEUC.349732
5. G. Zhao, X. Wang, J. Wang, and J. Liu, "Task recommendation for mobile crowd sensing system based on multi-view user dynamic behavior prediction," *Peer-to-Peer Netw. Appl.*, vol. 16, no. 3, pp. 1536-1550, 2023, doi: 10.1007/s12083-023-01504-x.
6. M. Schleicher, V. Unnikrishnan, R. Pryss, J. Schobel, W. Schlee, and M. Spiliopoulou, "Prediction meets time series with gaps: User clusters with specific usage behavior patterns," *Artif. Intell. Med.*, vol. 142, p. 102575, 2023, doi: 10.1016/j.artmed.2023.102575.