International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 4722-4731

https://doi.org/10.70517/ijhsa463404

Optimization of knowledge discovery methods for library data mining based on augmented learning

Xiuhua Wu[1] and Guoqiang Sang[2,*]
[1] Library (Archives), Zhejiang College of Security Technology, Wenzhou, Zhejiang, 325000, China
[2] School of Physical Education and Health, Wenzhou University, Wenzhou, Zhejiang, 325000, China
Corresponding authors: (e-mail: sdhzsgq1978@126.com).

Abstract Among the various services of smart libraries, knowledge discovery services are becoming more and more popular among users. Knowledge discovery based on data mining supports users to obtain convenient resources, as a result, this paper optimizes the library knowledge discovery method, utilizes the DDGP reinforcement learning algorithm for enhancement learning, and proposes a dual experience pool structure and hierarchical experience playback mechanism to improve the DDPG algorithm, and realizes the library accurate recommendation model based on enhancement learning. Experiments are carried out on different datasets, and the results of each evaluation index of the improved DDPG algorithm in this paper are better than those of the comparison methods, with the hit rate and the cumulative gain of the normalized discount improved by 0.132~0.380 and 0.074~0.308 compared with that of the DDPG algorithm, and the superior performance of the accuracy, the recall and the F-value are also obtained under different recommendation numbers. Experiments show that the library precision recommendation method based on augmented learning in this paper has better resource recommendation accuracy and can provide users with more personalized knowledge discovery services.

Index Terms augmented learning, data mining, knowledge discovery, recommendation model, DDGP, library

I. Introduction

With the development of the network information environment, digital libraries are faced with the challenges of information technology and the continuous development of users' information needs and rapid changes in the information competition environment [1]. On the one hand, with the rapid development of information technology, database technology, the "information explosion" phenomenon is becoming more and more obvious, the continuous upgrading of information storage methods (from databases to data warehouses, database clusters, etc.), are revealing that the data capacity shows an exponential increase in the speed of the speed [2]-[4]. However, the development of data processing technology is relatively backward, database technology is still stuck in the relatively simple entry, query, statistics, retrieval stage, the database of the data between the existence of relationships and rules, data group characteristics, data sets within the laws and trends, etc., but the lack of effective technical means to dig out, thus appearing the so-called "by data The so-called "drowned by data, but hungry for knowledge" phenomenon [5]-[8]. However, the level of people's demand for "knowledge" is escalating, and they are more and more dissatisfied with the information they obtain through simple inquiries. Instead, they want to obtain more useful and deeper "knowledge" to meet their learning and working needs [9]-[11]. Then it becomes a priority to find ways to extract the important information and knowledge hidden behind a large amount of data.

Knowledge Discovery (KDD) is a cross-discipline that emerged in the 1980s, which was developed in the context of deepening research in database technology, machine learning technology, statistical technology, artificial intelligence technology, etc. [12]. The rise of KDD is the result of a long history of research and development in data processing and analysis [13]. It has brought database technology to a more advanced stage, which is not only capable of querying and traversing past data, but also capable of identifying potential connections between data and discovering predictive, discrepant knowledge that is instructive [14], [15]. Bracco, A et al. discussed the potential of knowledge discovery through data mining (KDD) techniques that can extract new insights by identifying datasets [16]. Hammad, A and AbouRizk, S developed a five-step data based Knowledge Discovery (KDD) integration approach and converted existing multidimensional historical data into useful knowledge [17]. Molina-Coronado, B et al. provided a comprehensive overview of network intrusion detection methods from the perspective of Knowledge Discovery in Databases (KDD) process, discussing their effectiveness for use in data collection, preprocessing transformation, data mining and evaluation [18]. Shu, X and Ye, Y pointed out that knowledge discovery and data mining are emerging interdisciplinary fields that can complement traditional statistical methods by improving model fitting, revealing hidden patterns, identifying nonlinear effects and enriching scientific discovery [19]. These successful application examples have proved that the knowledge discovered by the use of knowledge discovery tools has a great guiding effect on production, scientific research and many other fields [20]. Therefore, knowledge discovery in libraries with huge information resources will certainly provide better knowledge guarantee for the development of human beings and the development and progress of society [21].

The study analyzes library knowledge discovery based on data mining from two aspects: data collection and data mining. Drawing on reinforcement learning theory, the DDGP algorithm is used to establish a library resource recommendation model with the user as the environment and the library recommendation system as the intelligent body. Aiming at the problem of inefficient optimization of the intelligent body when DDPG is applied to sparse reward recommendation scenarios, we propose a dual experience pool structure and a layered experience playback mechanism, which separates the storage space for zero- and non-zero- reward experiences, and controls the proportion of each type of data in the batch each time the intelligent body takes it by means of stratified sampling, so as to improve the utilization rate of the experience information, and thus enhance the reward density and long-term rewards. The school digital library data and Goodbooks-10k dataset are selected for experiments to explore the performance and enhancement of the proposed method for library resource recommendation and knowledge discovery services by comparing the hit rate and normalized discount cumulative gain of different models in the two dataset varieties, as well as comparing the model's accuracy, recall, and F-value with different recommendation numbers.

II. Library knowledge discovery technology applications

In recent years, with the establishment and improvement of the mechanism of sharing digital information resources, the physical link of resource access has been opened. How small and medium-sized libraries can utilize this rare opportunity to improve their service quality instead of simply existing themselves as nodes of this shared information network has become the focus of attention. Knowledge discovery based on data mining is one of the key technologies of smart libraries, providing a new source of power for innovative library services. From the application scenario of knowledge discovery, the key to the realization of knowledge discovery based on data mining lies in data and algorithms. The knowledge discovery system model is shown in Figure 1.

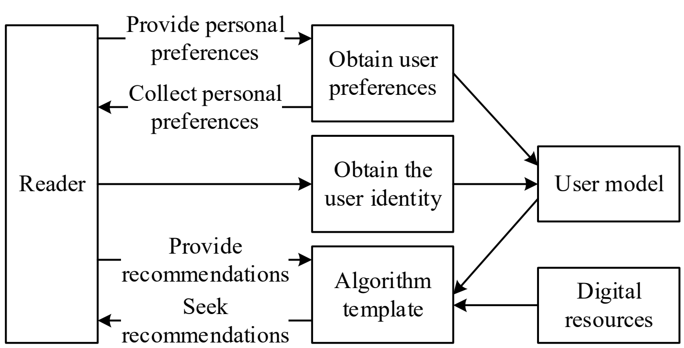

Figure 1: Model of knowledge Discovery System

II. A.Data acquisition

From the aspect of data, it mainly comes from two major data: reader information and digital resource library. The acquisition of reader information should be as comprehensive as possible, focusing on the collection of reader identification information and reader demand identification information. The reader's identification information can be obtained from the reader's library, while the reader's demand identification information can be obtained through various channels such as questionnaires, library micro-services, and search logs of the digital resource library. Combining the two, data can be extracted from reader discipline and reader need language description.

The data sources of digital repository should be as diversified as possible and the data format should be standardized. The main obstacle at present is the access to digital repository data. The major database vendors only open the data of limited fields, such as title, author, keywords, abstracts, etc., while for full-text data, they only provide data downloads in specific formats. Obviously, this practice only takes into account the readers' retrieval and utilization of primary literature, but not its secondary analysis and utilization. Such a format is not convenient for machine reading and program analysis, which increases the difficulty for data mining. As far as data mining applications are concerned, the whole data sample must be analyzed and mined in order to derive more effective data. As a joint purchaser of database co-purchasing and sharing, it should propose to the supply side to provide all machine-readable data, such as Unicode-encoded data. In addition, libraries themselves should promote the realization of multi-dimensional integration of digital resources and the formation of knowledge warehousing, and cannot simply rely on procurement alone to realize the construction of digital resources. The construction of digital resources in libraries should not only expand the sources of digital resources through the mechanism of co-build and share, but also endeavor to incorporate some free resources, such as traditional printed documents, oral history materials, original works on the Internet, and even social media information such as WeChat, Weibo, Tik Tok and other social media information, into the scope of knowledge warehousing.

II. B.Data mining

The difficulty of data mining lies mainly in the choice of algorithms. From the aspect of algorithms, the current application of data mining has been relatively mature: First, through a variety of data mining algorithms, libraries can boldly draw on the successful experience of the computer programming industry to encapsulate and packaged function function libraries, these algorithms are packaged, categorized according to the use of functions, and the establishment of template libraries, and at the same time, they can also be shared with their peers, according to the map, to achieve the needs of their respective functions. Secondly, through matching algorithms, the collected data are mined to form knowledge mapping secondary literature recommended for readers to read. Third, it supports readers to conveniently access high-quality knowledge resources based on problem scenarios (seeking recommendations), enabling them to carry out in-depth learning and creation, which is exactly where the value of data mining lies.

III. Augmented learning based accurate recommendation model for libraries

Library knowledge discovery service analyzes reader and article data through data collection and data mining, in order to provide readers with more personalized services, this paper uses augmented learning technology to optimize the library knowledge discovery method based on data mining, and proposes a recommendation algorithm based on reinforcement learning enhancement.

III. A. Enhanced learning

Reinforcement learning (RL) algorithms as a kind of machine learning algorithms, its training process is quite similar to the way of human learning, and thus it has attracted a lot of attention, and plays a very important role in the field of artificial intelligence path planning and other fields.

III. A. 1) Principles of Reinforcement Learning

Reinforcement learning mainly relies on the interaction between the intelligent body and the environment, the intelligent body is the main body that performs the task, it performs the strategy according to the state in the current environment and adjusts its value or strategy according to the rewards given by the environment after the execution of the strategy, and finally obtains the optimal action strategy in the process. The framework of reinforcement learning is shown in Figure 2.

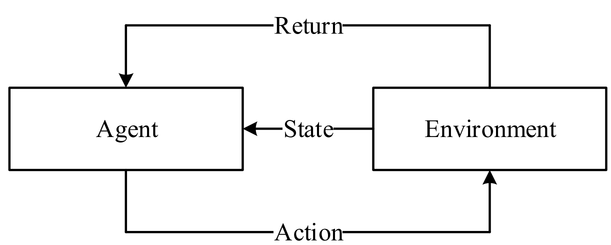

Figure 2: Framework of Reinforcement Learning

The essence of the main process of reinforcement learning is the Markov Decision Process (MDP), which is a mathematical model used to describe sequential decision problems. It is composed of five tuples $\langle S, A, P, R, \gamma \rangle$, the meaning of each of which is shown below:

S : represents the state space of the intelligent body, the state is a description of the environment, and the state will change after the intelligent body makes an action.

A : denotes the action space of an intelligent body, the action is a description of the behavior of the intelligent body and is an ensemble of results of the intelligent body's decisions.

P : denotes the state transfer probability function, $P(s_{t+1} | s_t, a_t)$ represents the probability that the intelligent body transfers to the state of s_t after executing the action a_t in the state of s_{t+1}.

R : denotes the reward function, $R(s,a)$ denotes the immediate reward obtained after performing action a in state s.

γ: denotes the degree of influence of the future reward on the current decision of the intelligence, usually denoted by γ, where $0 \le \gamma \le 1$.

The closer γ is to 1, the greater the influence of the future reward on the current decision.

The value function $V(s)$ is the expectation of the cumulative reward of the intelligent body after executing the action strategy π in the state of the current state s, all the way to the final state, which is denoted as $V_\pi(s)$, and its expression is as follows:

$$V_\pi(s) = E_\pi\left(\sum_{k=0}^{\infty} \gamma^k r_{t+k} \mid s_t = s\right) \tag{1}$$

The discount factor γ is an important hyperparameter. Given an immediate reward r_{i+k} and an environmental state s_t at any moment t, for any action strategy π, the value function of the moment t system in state s_t is:

$$V_\pi(s_t) = r_t + \gamma \sum_{s_t \in S} P(s_{t+1} | s_t, a_t) V_\pi(s_{t+1}) \tag{2}$$

where r_i represents immediate reward and $P(s_{t+1} | s_t, a_t)$ represents the probability of transferring from the current state to the next state.

The action strategy π defines the behavioral choices of the intelligence in different states. Specifically, $S \to A$ is a mapping from the state set S to the action set A. The core goal of reinforcement learning is to find the optimal action policy π^* that maximizes the cumulative reward of the intelligent body, i.e:

$$V_{\pi^*}(s) = \max_{a \in A}\left\{ r_t + \gamma \sum_{s_{t+1} \in S} P(s_{t+1} | s_t, a_t) V_\pi(s_{t+1}) \right\} \tag{3}$$

III. A. 2) Deep deterministic strategy gradient

Deep Deterministic Policy Gradient (DDPG) is a reinforcement learning algorithm that can effectively deal with a continuous action space and make optimal decisions based on environmental information.

In DDPG, the exploration and learning of action strategies are two different processes, with action exploration using a stochastic strategy and learning strategy based on a deterministic strategy. It continues to use the experience pool reuse technique of the DQN algorithm to minimize the correlation between samples and introduces the Actor-Critic architecture, which employs two neural networks, Actor and Critic, to estimate the optimal policy and the state value function, where the Critic network exploits the difference between the current state and the next state to update the value network. TD error to update the weights of the value network, and the Actor changes the execution probability of all selectable actions in the current state by the evaluation given by Critic.

The DDPG algorithm collects sample data (s_t, a_t, r_t, s_{t+1}) by interacting with the environment and stores it in the experience pool. In the sample data of this paper, s_t represents the state information observed by the ship agent from the planning environment at the time of t, a_t is the action strategy executed by the ship agent based on the observed state information s_t, s_{t+1} represents the new state information observed by the ship agent from the environment after executing the action strategy a_t, and r_i The reward value for the ship agent for performing an action a_t in the state s_t. In learning the strategy for ship collision avoidance path planning, the algorithm randomly draws sample data from experience leaks for learning.The DDPG algorithm involves the updating of four neural networks during the learning process in the following manner:

The current Actor network update strategy gradient is:

$$\nabla J(\theta) \approx \frac{1}{N} \sum_i \nabla a Q(s, a \mid \theta^Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} \mu(s \mid \theta^\mu)|_{s_i} \tag{4}$$

where $J(\theta)$ is the objective function with respect to the Actor network parameter θ, θ^Q and θ^μ denote the parameters of the Critic network and Acton network respectively, s_i is the current state, a is the current Actor network output action, and $\mu(s|\theta^\mu)$ is the output action of the Actor network. The current state and action are first evaluated by the Q function of the Critic network to compute $\nabla aQ(s,a|\theta^Q)$. This gradient is then used to multiply the policy gradient $\nabla_{\theta^\mu}\mu(s|\theta^\mu)$ of the Actor network. The policy gradient represents the direction of the gradient of the Actor network's output actions in the current state, i.e., the direction of adjustment for actions that maximize the cumulative reward.

The current Critic network is updated by minimizing the loss function, which is:

$$L(\theta^Q)=\frac{1}{N}\sum_i(y_i-Q(s_i,a_i|\theta^Q))^2 \tag{5}$$

where the target value is calculated by the formula:

$$y_i=r_i+\gamma Q'(s_{i+1},\mu(s_{i+1}|\theta^{\mu'})|\theta^{Q'}) \tag{6}$$

where $L(\theta^Q)$ denotes the loss function with respect to the parameters of the Critic network θ^Q, N denotes the number of samples, r_i denotes the immediate reward at the current moment, s_{i+1} denotes the next state, $\theta^{Q'}$ denotes the parameters of the target Critic network, $Q(s_i,a_i|\theta^Q)$ denotes the Q-valued function of the Critic network, and $\mu(s_{i+1}|\theta^{\mu'})$ denotes the target Actor network given the the action of the next state. The parameters of the Critic network are updated by minimizing the gap between the target Q value and the current estimated Q value.

Unlike DQN which updates the target network parameters by directly copying the weights, DDPG uses a soft update to update the target network parameters. Its update method is as follows:

$$\begin{cases}\theta^{Q'}=\tau\theta^Q+(1-\tau)\theta^Q\\ \theta^{\mu'}=\tau\theta^\mu+(1-\tau)\theta^{\mu'}\end{cases} \tag{7}$$

The DDPG algorithm is a deterministic policy algorithm, which, although algorithmically efficient, suffers from a lack of exploration capacity. In order to explore the action state space more comprehensively, the DDPG algorithm introduces an exploration quantity during training. Specifically, it adds this exploration quantity to the current Actor network output action, making the action executed by the intelligent body change from a deterministic value to a random value. Then, the action values are sampled from this random process, and the final action executed by the intelligent body is shown in equation (8):

$$a_t=\mu(s_t|\theta^\mu)+\lambda \tag{8}$$

where λ is the exploration quantity, a random number obeying a Gaussian distribution.

III. B. Recommendation algorithms based on augmented learning

III. B. 1) Holistic thinking

In a real online recommendation scenario, the interaction process between the library recommendation system and the user is very similar to the interaction process between the intelligent body and the environment in reinforcement learning. When the reinforcement learning theory is applied to the recommendation scenario, the user is taken as the environment and the recommendation system is taken as the intelligent body, and the whole process is modeled as follows: the recommendation system senses the user's state and makes recommendations according to a certain strategy, the user gives feedback (rewards) according to the recommendation results and changes its own state, and the intelligent body updates its own parameters through the observation of the relationship between the state, the action and the rewards, in order to expect to get the maximization of the long-term rewards. Therefore, compared to applying common statistical optimization methods in recommendation scenarios, reinforcement learning theory helps maximize long-term marketing gains when applied.

In the research on deep reinforcement learning algorithms, the DDPG algorithm greatly reduces the algorithm's need for the amount of experience by applying deterministic policy gradients, which improves the optimization efficiency, so this paper is mainly based on the DDPG algorithm to construct an accurate recommendation model for libraries.

III. B. 2) DDPG Algorithm Improvement

In some recommendation scenarios, where the user mainly gives feedback in the form of clicks and in most cases there is no click action, the rewards fed back to the intelligent body are zero-valued, which belongs to sparse reward scenarios. When the DDPG algorithm is directly applied to the recommendation in this scenario, it will lead to inaccurate gradient direction, which will seriously reduce the optimization efficiency of the intelligent body, and then lead to further decrease of reward density, thus causing a vicious circle. To address this problem, this paper proposes a dual experience pool structure with a hierarchical experience playback mechanism to optimize the DDPG algorithm. The overall flow of the algorithm is as follows:

First, the process from inputting user ID, history product ID and candidate product ID to generating the matching score can build a personalized recommendation model based on deep learning, and this process can apply the idea of multifarious features and multifarious user portraits, i.e., fusion of user and product content features and hidden factor features, and fusion of user's personal representation and user's interest representation. The offline training means can be pre-trained to obtain the multivariate feature representation of the historical record products, the multivariate feature representation of the candidate set products, the multivariate feature representation of the users and the parameters of the Actor network.

Next, the user history multivariate product feature representation and the user multivariate feature representation are obtained as state vectors s_t by the state encoding module. The Actor network then takes s_t as input to get the strategy vector a_t and sums it with the noise vector n_t according to a certain exploration probability e to get the final a_t. where n_t is a vector of the same shape as a_t and each of its elements is a random variable obeying the standard normal distribution N(0, 1). This is done to balance the exploration of unknown strategies with the utilization of known strategies. After this, a_t is operated with the candidate product feature representations in an inner product operation to obtain the matching score of each candidate product, which further ranks the candidate products to generate personalized recommendation results to the user. Finally, the user makes a choice for this recommended product and also feeds a real-valued reward r back to the intelligent body, which changes the set of products in the user's history and thus affects the state of that user at the next moment of interaction s_{t+1}. At this point, the user and the deep reinforcement learning intelligent have completed a full interaction process.

Whenever the user and the intelligent body complete a round of interaction, the intelligent body gets a set of experience information (s_t, a_t, r_t, s_{t+1}), and then chooses to store it in Memory0 or Memoryl according to whether r_t is a zero-value or not (the user clicks or not). A certain amount of empirical information is taken from Memory0 and Memoryl to form a data batch according to a set percentage of empirical playback for each training, and the network parameters are updated once by applying the DDPG algorithm.

IV. Experimental results and performance analysis

IV. A. Data sets

In the experiments, this paper uses two real datasets: a school digital library data and Goodbooks-10k dataset to evaluate the model proposed in this paper. The school digital library data contains 504,245 borrowing records. The Goodbooks-10k dataset contains a book recommendation dataset with 912,705 borrowing records.

Since some users have borrowed only a small number of books, a small amount of data cannot form a sequence, so the data of users who have borrowed more than or equal to 4 are used. Most of the sequences in the two data sets are below 20 in length, while the number of books is above 10,000, which also indicates the sparseness of the book data. In the process of training and testing, the last item of these sequences is used as the test data and the others are used as the training data to get the final results.

IV. B. Comparison of experimental methods

To demonstrate the superiority of the model proposed in this paper, the model is compared with the following models:

CF: A collaborative filtering algorithm is an algorithm that utilizes a set of preferences with similar interests and common experiences to recommend information of interest to the user.FISM: is an item-item collaborative filtering algorithm but does not use the attention mechanism to distinguish the weights of historical data.NAIS: is an item-item collaborative filtering algorithm that uses the attention mechanism to distinguish the weights of historical data. It is used as a basic recommendation model in this paper.DDPG: is a reinforcement learning algorithm that can effectively deal with continuous action space and make optimal decisions based on user feature information.HRL: is an algorithm that is jointly trained using a basic recommendation model and a hierarchical reinforcement model.

In the experiments, accuracy, recall, F-value, as well as the hit rate of the first K items (HR@K) and the normalized discount cumulative gain of the first K items (NDCG@K) are used as evaluation metrics. Specifically, HR@K is a recall-based metric that measures the percentage of successful recommendations of instances in the top K items, and NDCG@K is an accuracy-based metric that indicates the predicted position of instances. In this paper, K is set to 5 and 10, and all metrics including 1 positive and 99 negative instances are computed to obtain the average score of all user sequences.

IV. C. Analysis of experimental results

The experimental results for the two datasets are shown in Figures 3 and 4. For school digital library data, the improved DDPG algorithm of this paper outperforms other methods in terms of knowledge discovery performance of library data, with the results of HR@5, HR@10, NDCG@5, NDCG@10 of 0.828, 0.875, 0.625, 0.668, which is an improvement compared with the DDPG algorithm by 0.380, 0.291, 0.308 and 0.284. For Goodbooks-10k data, the results of HR@5, HR@10, NDCG@5, NDCG@10 of this paper's model are 0.608, 0.719, 0.371, 0.443, which are significantly better than the other models in terms of recommendation results. Its recommendation hit rate and normalized discount cumulative gain for library data are improved by 0.184, 0.132, 0.074, and 0.082 over the DDPG algorithm before enhanced learning.

The two algorithms, FISM and NAIS, are item-item collaborative filtering algorithms. They use a deep learning approach, and the NAIS algorithm adds an attention mechanism to the FISM algorithm, but sparse data can have a large impact on the training process of these two algorithms, leading to poorer final results.The HRL algorithm obtains slightly better results, but they are not particularly perfect.

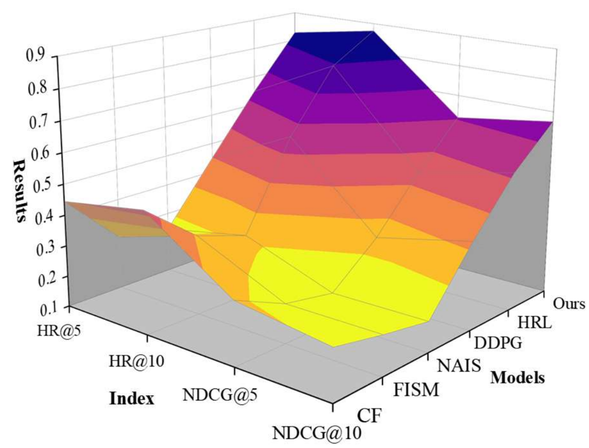

Figure 3: Comparison of experimental results of school digital library data

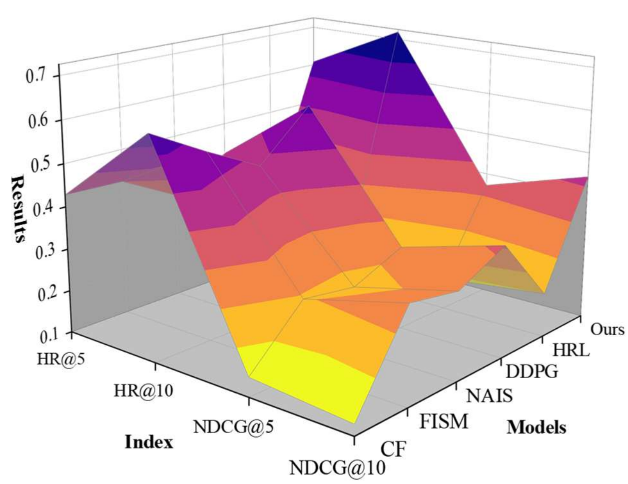

Figure 4: Comparison of experimental results of Goodbooks-10k data

In order to verify the accuracy of the recommendation results, the accuracy rate, the recall rate and the composite value F, which are commonly used to evaluate the effectiveness of recommendations, are used to evaluate the recommendation results. Among them, the accuracy rate is the proportion of the number of hit recommendations to the total number of recommendations, the recall rate is the proportion of the number of recommended hit items to the theoretical number of recommended hits, and the comprehensive index F value is the weighted reconciliation average of the accuracy rate and the recall rate. In this paper, the DDPG recommendation model improved based on augmented learning is used to compare with CF, FISM, NAIS, DDPG and HRL algorithms to compute the accuracy rate, recall rate and F value of the comprehensive indexes under different number of recommendations, respectively, and the results of the comparison of the accuracy rate, recall rate and F value under different number of recommendations are shown in Fig. 5, Fig. 6 and Fig. 7. The results show that when the number of recommendations rises from 5 to 30, the accuracy of the recommendations of the 6 methods decreases sequentially, while the recall and F-value gradually increase. When the number of recommendations is the same, the recommendation algorithm with improved DDPG has higher accuracy, recall, and F-value than the other five recommendation algorithms, and the recommendation effect is better. Under different library recommended titles, the average accuracy, recall and comprehensive index F-value of the improved DDPG recommendation algorithm are 0.715, 0.521 and 0.599, respectively, while the average accuracy, recall and comprehensive index F-value of the DDPG algorithm are 0.612, 0.347 and 0.437, which are significantly lower than that of the improved DDPG recommendation model, indicating that this paper utilizes the reinforcement learning method to enhance the effectiveness of the library precision recommendation model.

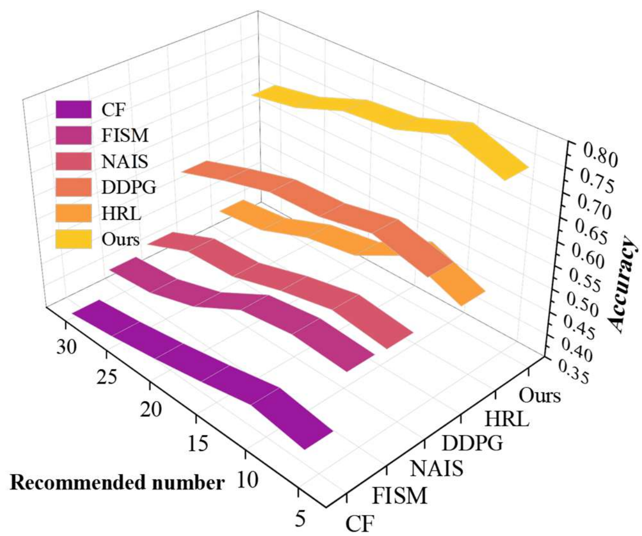

Figure 5: The accuracy comparison of different recommendation numbers

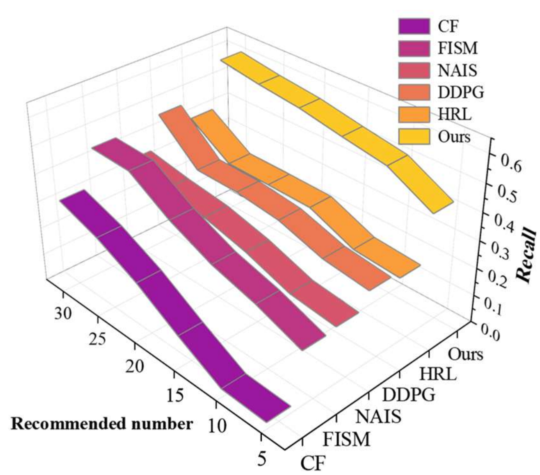

Figure 6: The recall comparison of different recommendation numbers

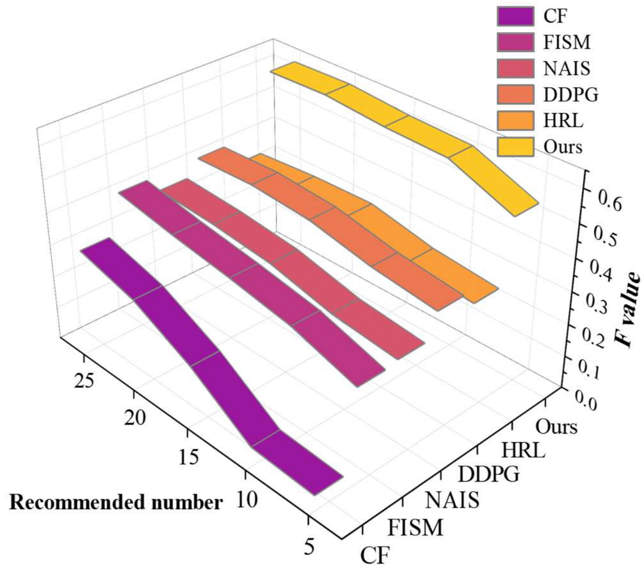

Figure 7: The F value comparison of different recommendation numbers

V. Conclusion

Smart service is an important content of smart library construction and the way of survival for future libraries, in which knowledge discovery based on data mining is the key technology to realize smart service. In order to optimize the knowledge discovery service of libraries, this paper uses reinforcement learning algorithm to enhance the performance of library data recommendation, proposes to improve the accurate recommendation model of DDGP, and conducts experiments to analyze its recommendation effect.

Through experiments on the school digital library dataset and the Goodbooks-10k dataset, it is found that the hit rate and normalized discount cumulative gain results of the improved DDPG algorithm are better than those of other algorithms, and its HR@5 and HR@10 on the two datasets are 0.132~0.380 higher than that of the DDPG algorithm, and the NDCG@5 and NDCG@10 are 0.074~0.308 higher than those of the DDPG algorithm. At the same time, the average accuracy, recall rate and F-value of the comprehensive index were 0.715, 0.521 and 0.599 under the number of recommended data of different libraries, which were higher than those of the other five comparison methods. Experiments show that the algorithm achieves good results in improving the recommendation accuracy.

There are still many directions worth exploring for augmented learning in library recommendation systems, such as considering the long-term interests of users, introducing causal inference, and exploring more efficient inference mechanisms. With the deepening of artificial intelligence research, it is believed that augmented learning will certainly play a greater role in intelligent information systems to provide users with more accurate, efficient and personalized recommendation services. Therefore, university libraries should start from knowledge discovery application scenarios, efficiently utilize the existing mature data mining methods and technologies to solve the various problems faced by readers, improve the quality of library services, and play the role of libraries in knowledge dissemination.

Funding

This work was supported by The National Social Science Fund of China (Project Number: 24BTY044).

References

[1] Yaqin, M. A. (2022). Strategy of library development towards digital library. Khatulistiwa: Jurnal Pendidikan Dan Sosial Humaniora, 2(2), 52-69.
[2] Roy, S. G., Sutradhar, B., & Das, P. P. (2017). Large-scale metadata harvesting—tools, techniques and challenges: A case study of National Digital Library (NDL). World Digital Libraries-An international journal, 10(1), 1-10.
[3] Jha, S. K. (2023). Application of artificial intelligence in libraries and information centers services: prospects and challenges. Library hi tech news, 40(7), 1-5.
[4] Dong, N. (2024, March). Research on Knowledge Discovery Service in Digital Libraries Based on Deep Learning. In 2024 13th International Conference on Educational and Information Technology (ICEIT) (pp. 328-333). IEEE.
[5] Weng, C. H. (2016). Knowledge discovery of digital library subscription by RFC itemsets. The Electronic Library, 34(5), 772-788.
[6] Sun, Z., & Stranieri, A. (2020). A knowledge discovery in the digital age. PNG UoT BAIS, 5(1), 1-11.

[7] Marcos-Pablos, S., & García-Peñalvo, F. J. (2020). Information retrieval methodology for aiding scientific database search. Soft Computing, 24(8), 5551-5560.
[8] Pawar, S. S., Manepatil, A., Kadam, A., & Jagtap, P. (2016, March). Keyword search in information retrieval and relational database system: Two class view. In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) (pp. 4534-4540). IEEE.
[9] Gan, W., Lin, J. C. W., Chao, H. C., & Zhan, J. (2017). Data mining in distributed environment: a survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 7(6), e1216.
[10] Bhat, P., & Malaganve, P. (2023). Design and Implementation of Social Media Mining–Knowledge Discovery Methods for Effective Digital Marketing Strategies. Data Engineering and Data Science: Concepts and Applications, 23-48.
[11] Dhiman, A. K. (2011). Knowledge discovery in databases and libraries. DESIDOC Journal of Library & Information Technology, 31(6).
[12] Ho, T. B. (2016). Knowledge discovery. Knowledge Science, 70-93.
[13] Nwagu, C. K., Omankwu, O. C., & Inyiama, H. (2017). Knowledge Discovery in Databases (KDD): an overview. Int J Comput Sci Inf Secur (IJCSIS), 15(12), 13-16.
[14] Rahman, F. A., Desa, M. I., Wibowo, A., & Haris, N. A. (2014). Knowledge discovery database (KDD)-data mining application in transportation. Proceeding of the Electrical Engineering Computer Science and Informatics, 1(1), 116-119.
[15] Ghazal, M. M., & Hammad, A. (2022). Application of knowledge discovery in database (KDD) techniques in cost overrun of construction projects. International Journal of Construction Management, 22(9), 1632-1646.
[16] Bracco, A., Falasca, F., Nenes, A., Fountalis, I., & Dovrolis, C. (2018). Advancing climate science with knowledge-discovery through data mining. npj Climate and Atmospheric Science, 1(1), 20174.
[17] Hammad, A., & AbouRizk, S. (2014). Knowledge discovery in data: A case study. Journal of Computer and Communications, 2014.
[18] Molina-Coronado, B., Mori, U., Mendiburu, A., & Miguel-Alonso, J. (2020). Survey of network intrusion detection methods from the perspective of the knowledge discovery in databases process. IEEE Transactions on Network and Service Management, 17(4), 2451-2479.
[19] Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. Social Science Research, 110, 102817.
[20] Cios, K. J., Pedrycz, W., & Swiniarski, R. W. (2012). Data mining methods for knowledge discovery (Vol. 458). Springer Science & Business Media.
[21] Dumouchel, B., & Demaine, J. (2006). Knowledge Discovery in the Digital Library: access tools for mining science. Information Services and Use, 26(1), 39-44.