Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 4, pp. 2589–2595, DOI 10.12694/scpe.v25i4.2883

RESEARCH ON THE APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGY IN THE BANKING INTERNET FINANCE INDUSTRY

TIANHAO ZHANG*

Abstract. This paper presents a collaborative filtering algorithm based on reinforcement learning theory. Then, the personalized bank financial recommendation system for users is constructed in the massive data environment. Tags mimic different types of user interest points to build a representative personalized data set. The collaborative screening of bank financial products is realized using the simulation results and users' historical access records. The ranking calculation of related financial products is added to the general bank financial product recommendation system. This method can more accurately express the query results for a specific user. It is found that the collaborative filtering algorithm based on enhanced learning theory can improve the efficiency of collaborative screening of bank financial products. The best results can be obtained by combining the two organically. This paper proposes that the recommendation algorithm of reinforcement learning bank financial products based on user preference and collaborative filtering is feasible.

Key words: Bank financial products; Reinforcement learning; Feature extraction; Vector space; Adjacency matrix; Collaborative filtering; Artificial intelligence

1. Introduction. Under the background of business optimization and industry reform in the financial industry, Chinese banks and various financial organizations have developed various financial products according to market needs to meet the needs of various customers. However, due to the increasing number and types of financial products, the quality and evaluation of various financial products have also appeared uneven. Customers often spend a lot of energy searching for the goods they want, affecting their shopping experience and significantly impacting the bank's operating income. Customers can quickly and accurately search and select the goods they need in many financial commodities, an essential problem the banking industry needs to solve.

Scholars use the recommendation system as a means to analyze users' consumption habits and interests in the past to discover the potential preferences and needs of customers. In this way, the customer is presented with the most suitable product information. Literature [1] proposes a model based on the interaction between users and products. Compared with other content and mixed recommendation methods, combined filtering recommendation is convenient and fast and does not require additional business knowledge. It can assist users in exploring new commodity needs and is more suitable for the e-commerce environment. However, this method faces the problems of data sparsity and cold start, which reduces the recommendation accuracy. Literature [2] introduces features and evaluation information such as user and commodity attributes. It can solve problems such as sparse data and cold start to improve the product's recommendation performance. User preference is a data analysis method that focuses on customers and products. The historical interaction information of customers is used as a link to establish the relationship between customers and products. Use the differentiated semantic construction to explore the potential needs of customers. This improves the recommendation system's accuracy, variety, and comprehensibility [3]. In this paper, the bank financial product recommendation system is established based on user preferences and collaborative filtering reinforcement learning.

2. Design of bank financial product recommendation application system under multi-source big data environment. A personalized visualization solution for big data is built based on the Hadoop processing platform and MapReduce computing architecture [4]. The system's primary functions include data acquisition, fusion, algorithm realization and business service (fig. 2.1).

*Faculty of Business and Economics, Monash University, 3800, Wellington Road, VIC, Australia (tzha0020@monash.student.edu)

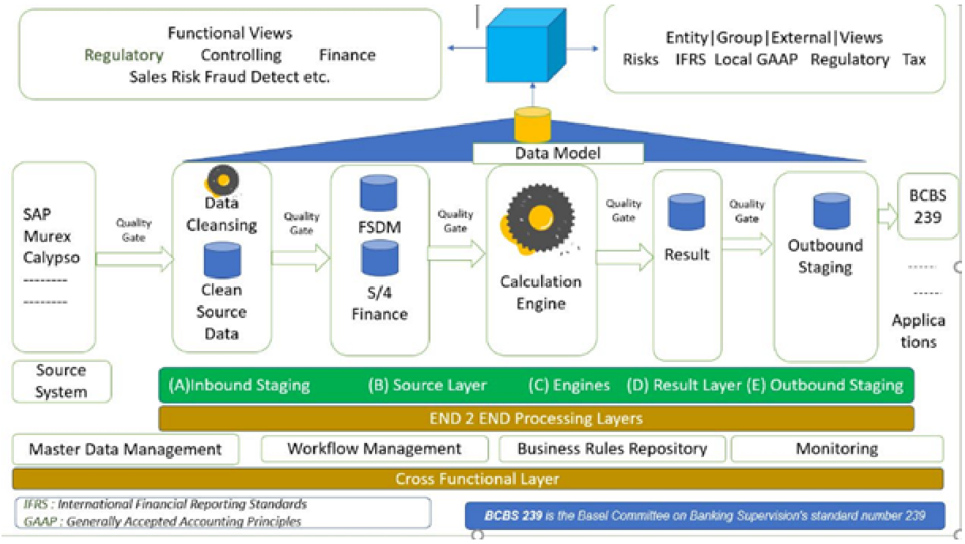

Fig. 2.1: *Architecture of bank financial product recommendation system for multi-source data.*

2.1. Data Collection. The low-cost HDFS distributed file system can process and update data from terminals such as PCS, mobile, cloud, and sensors in real-time. HDFS includes a file structure that exists in the form of inheritance. The different file systems are very similar [5]. Users can save files to their created folder or transfer files to another folder and rename them. The DFS Shell interface in HDFS enables users to access its data. This high throughput feature allows easy access to programs with large data sets.

2.2. Data Fusion. The MapReduce algorithm is used to merge the collected multi-source data. Erase user data, filter noise, and remove excess information. The standardization process is carried out to make the system's data structure get a unified specification [6]. Through the interactive analysis of the characteristics of each data to complete the transformation and synthesis of information. MapReduce can create a different number of nodes by using a standard server. It can realize the automatic segmentation and processing of the operational data in the cluster through the system of data positioning, fault tolerance optimization and other small and tedious content in the calculation work. This makes the system developer's job easy. According to the Lisp programming idea, a simple operation interface is established. Integrate Map and Reduce to complete large-scale data-oriented programming and parallel computing.

2.3. Algorithm implementation. This project intends to use Hadoop as a platform and combine visual analysis technology to discover user data from multiple sources. A large-scale data mining method based on attributes is proposed. Big data mining algorithms mainly include six types: 1) They can combine various data items into corresponding categories. It is mainly used in data classification, preference prediction and so on. 2) The classification problem faced by clustering is uncertain. The correlation between different categories of data is minimal, but the correlation between the same type of data is very high. 3) The degree of correlation between the variables was obtained through multiple linear regression of the data. It can be used in the research of forecasting, error control and so on. 4) The result of association rules is to mine the interaction between data items. From a single piece of data, infer the implied goal of the association. It is often used to predict user requirements. 5) Neural networks are artificial intelligence technologies that mimic human thinking. It is characterized by a network system composed of multiple neurons, which can independently self-process and decentralize data storage. It has higher advantages in learning and promotion. 6) Web data mining uses mining methods from the vast network data to find hidden valuable information and patterns. Then, it comprehensively analyzes and processes the structure and behavior of web pages. The main advantages of this method are high parallelism and real-time solid dynamics.

2.4. Application Services. The front-end application of the platform is the personalized service of bank financial products. It realizes personalized recommendations, push, search and other functions. A personalized recommendation is based on the collected information for its corresponding opinions. Help customers find the right product and make a decision [7]. The content of the suggestion should be consistent with the user's immediate situation. It can accurately sense the user's movements. It is both novel and timely. For example, if a user has recently and frequently viewed skin care products on the site, the recommendation function can be implemented based on the user's age, income and past spending habits. Personalized push refers to using mobile platforms to recommend potential financial products to consumers. Personalized search means that users can obtain the exact needs in the search results and present the relevant knowledge content to users.

3. Collaborative filtering algorithm of reinforcement learning. First, establish the user interest vector associated with the tag. $\lambda_t \Theta_c^T$ represents the value function of the user c at time $t.\lambda_t$ is the proportion of different labels in the user c. It represents the area that each tag occupies in the user. Θ_c is the ratio of the frequency of each tag occurring in a user c to that tag occurring in many users [8]. It reflects the significant differences of individual marker vectors λ_t in the population. The results show that the influence of users on the evaluation results of network resources decreases with the development of the network. It tracks the user's most recent interest vector. A method based on linear approximation of function is proposed to study the user's interest vector. The following formula can express the reinforcement learning weighted correction algorithm:

$$\Lambda_{i+1,k} = \Lambda_{i,k} - \frac{1}{2}\eta\mu^{n-i}\frac{M_k}{\sum_{j=1} M_j}\nabla_\lambda\left[\varphi^\beta(c) - \varphi_i(c)\right]$$

$$\nabla_\lambda\left[\varphi^\beta(c) - \varphi_i(c)\right]^2 - 2\left(S_{i+1} + \delta\hat{Y}\left(c_{i+1}\right) - \hat{Y}\left(c_i\right)\right)\nabla_\lambda\hat{Y}_i(c)$$

So, formula (1) can be transformed into

$$\Lambda_{i+1,k} = \Lambda_{i,k} + \eta\mu^{n-i}\frac{M_k}{\sum_{j=1} M_j}\left(S_{i+1} + \delta\hat{Y}\left(c_{i+1}\right) - \hat{Y}\left(c_i\right)\right)\nabla_\lambda\hat{Y}_i(c)$$

$\Lambda_{i+1,k}$ is the weight that the $i+1$ dimensional tag has for the Step user $i+1$ after accessing the data. M_k is the number of k dimensional tokens that occur for that user. $\sum_{j=1}^n M_j$ shows the number of displays of all tags in the n dimension [9]. Where S_{i+1} is the value of the resource assessed by the user at time $i+1$. The value function of user c at the i time is

$$\hat{Y}_i(c) = \Lambda_i\Theta_S^T$$

So $\nabla_\lambda\hat{Y}(c) = \Theta_c^T$. Here, each dimension of Θ_c under the Laplacian correction is represented as follows

$$\psi_c\left(x_j\right) = \frac{1 + x_{c,j}}{\sum_{c=1}^N 1 + x_{c,j}}$$

$x_{c,j}$ is the number of j dimensional tabs displayed for the user $c.N$ is the number of users. Resources with different access times are graded using μ. The gradient learning algorithm extracts the intrinsic weights of representative eigenvalues [10]. The current vectors of interest are affected by the two factors λ_4 and Θ_c. Calculate the distance between the user group that has historically accessed the predicted resource and the current user. Use K users closest to the current user to estimate the value of future users.

4. TDIDF and text classification. 4. TDIDF and text classification. TFIDF is often used to calculate text weights in vector space models. In text recognition, domain weighting functions involve two fundamental learning modes: K-nearest neighbor and support vector machine. The components were weighted by the TFIDF method. z particular research area z may consist of a vector s of the individual files it contains. Suppose s represents the vector of file $s'.s'$ and the corresponding category vector z have a cosine distance [11]. It represents how close a file is to a particular category. It represents how similar the document is to a particular class. The value is the largest. It is the class z to which s' belongs.

4.1. PageRank. After the corresponding search results are obtained through the query, they must be appropriately calculated to be sorted according to the degree of importance. One of the best ways to do this is to use the web page's links. Google is an example of this [12]. This method uses the link operation method proposed by Brin and Page. If you think of A Web page and its connections as a directed graph $R = Q($ Page,Link $)$. This can be illustrated by the adjacent matrix Q. The element of Q can be defined as

$$Q_{ij} = \begin{cases} 1, \text{ Link } i \rightarrow \text{ jexists} \\ 0, \text{ Otherwise} \end{cases}$$

$i, j \in (1 \text{ L} n), n$ represents the number of web pages. If the probability of starting from one page to another is 1 , the elements of the Q matrix are operated with $Q_{ij} = Q_{ij} / \deg(i)$. It can represent the possibility of sending from one web page to another. $\deg(i)$ refers to the page outlet point. Q is called an arbitrary matrix of lines [13]. From the characteristics of Web page connection, Q and Markov chain are consistent. Modify the Q matrix. Replace all rows of zeros in the matrix with $u = \left(\frac{1}{n}\right) e^T \cdot n$ is the number of web pages. e^T is the vector of all 1 line. Q is changed to $Q' = Q + s \cdot u^T$. The $s = \begin{cases} 1, \text{ if } \deg(i) = 0 \\ 0, \text{ otherwise} \end{cases}$ here is called the jump page indicator [14]. It proves that Q is a random matrix of columns. Q is an arbitrary transfer matrix for R. So, PageRank can be used as the limit value for the following recursive procedure

$$a_j^{(t+1)} = \sum_i Q'_{ij} a_i^{(t)} = \sum_{i \rightarrow j} a_i^{(t)} / \deg(i)$$

The above formula can transform the cost vector to find Q'. Since A contains an element of o, it does not ensure that Q' is the largest eigenvector. The reason for this is that Q' can be simplified. Q' can include several subgraphs that are disconnected from each other.

$$W = zQ + (1 - z) e u^T, e = (1, 1, 1 \text{ K}, 1)^T$$

$z \in (0, 1)$. In most literature it is set as $z \in [0.85, 1).z$ is known as the Ding ling coefficient. Processing W allows $W_{ii}^{(t)} > 0(i, t \in (1 \ldots n))$ to be tested. So W is not periodic. We can think of Q' as a non-periodic irreducible positive row type random matrix. The convergence of $a^{(t+1)} = W^T a^{(t)}$ in some respects is obtained from the Perlen-Frobenis law. Its maximum intrinsic property vector is positive.

4.2. Re-calculation of the value of the website.

4.2.1. Keyword extraction. In the system covered in this paper, it is necessary to extract domain-specific keywords. Some scholars have proposed a lexical weighting method to classify specific categories. This formula ensures that the weight of certain irrelevant words is low. This ensures that the intrinsic properties of the eigencharacters are fully reflected in the classification [15]. This article uses a top-n keyword method called option value. The selected file source is a separate domain file and has been manually categorized. Each key word is weighted. Suppose that the particular class is represented by the vector $S_i = \{(t_j, \lambda_j), j \in (1, m)\}, i \in (1, n)$. Here n represents the total number of classes. m represents the number of class keywords in a particular class. (t_j, λ_j) represents the specific keyword that has its weight. File $s_i : s_i = \{(t_j, \lambda_j), j \in (1, m)\}, i \in (1, n)$ is represented by vector $s_i.n$ stands for the number of files. m indicates the number of keywords in the file. (t_j, λ_j) indicates the weight of the keyword in the file.

4.2.2. Download Link. The user access sequence of the system is obtained through the analysis of server login records. This sequence lets you download the user's visited page from the Internet (Figure 4.1).

4.2.3. Construction of web connection matrix. This system has two kinds of web connection: the host link and the primary link. The two are linked in different proportions. Set the weight of the internal link to 3/4. Link weight between hosts 1/4. Then the connection matrix Q is represented by

$$Q_{ij} = \begin{cases} \frac{3}{4 * / \deg(\text{ in tra })}, i, j \in G_m \\ \frac{1}{4 * / \deg(\text{ in tra })}, i \in G_m, i \in G_n, m \neq n 0, \text{ O therwise} \end{cases}$$

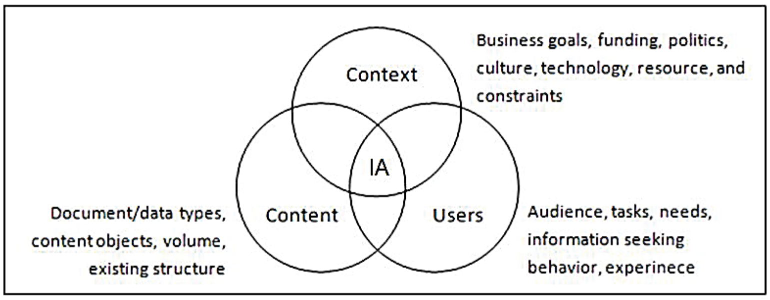

Fig. 4.1: *Hierarchical information structure of a Web site.*

Table 5.1: *Experimental data set.*

Data set	User number	Test the number of financial products	Number of labels	Average number of user tests	Average number of user tags	Average label users
1	786	2976	16663	3.938	22.063	0.049
2	779	2985	16663	3.990	22.271	0.048
3	808	3093	16663	3.979	21.469	0.050
4	807	3145	16663	4.052	21.500	0.050
5	807	2963	16663	3.917	22.063	0.049

4.2.4. PageRank calculation. The following data are stored in the local database: 1) The pages downloaded from the web page according to the login history. 2) Use categories to describe the vector of the web page. 3) Associate the link information of the page and keep two sequences: the sequence of the page is arranged according to the cosine distance of the page to the category. The PageRank formula calculates it. The two queues may consider using different weights to determine the importance of files in a particular category.

$$\operatorname{score}(s) = \eta^* \operatorname{sim}(s, S) + (1 - \eta) * PR(s)$$

s is used to store partial pages. $\operatorname{sim}(s, S)$ is the cosine from class D to class s. $PR(s)$ is the level of the page on page $s.\eta \in (0, 1)$ is the weighting factor. The result is a new sequence. It represents s ranking of pages from high to low importance for a particular S level.

5. Experiment.

5.1. Test data set. The empirical data comes from the customer data of Santander Bank. Santander Bank is the second-largest bank in Europe and the largest bank in Spain. The bank generates a large amount of customer behavior data monthly and continues expanding new financial products. It includes the user's number, gender, age, tag data, type of financial product purchased, customer level, etc. There are 74,549 users and 11,126 financial products. Among them, 10,416,723 people purchased financial products, including 99,563 marks. Five samples were drawn from the sample set to ensure the stability of the experiment. Each data set is selected from 1,000 randomly selected users, and each set of data is selected from 1,000 randomly selected users who have purchased at least three financial products. Table 5.1 lists the characteristics of the data.

5.2. Evaluation criteria. The training samples are evaluated by the actual scores of each user on the training set and the predicted results. The difference between the score and the actual score is used as the evaluation criterion to predict the selected M financial products. The mean square error of evaluating system

Table 5.2: *Experimental results.*

Algorithm	**1**	**2**	**3**	**4**	**5**	**Avg**
random	2.402	2.343	2.370	2.331	2.300	2.349
user-based	1.285	1.314	1.247	1.329	1.367	1.308
item-based	1.422	1.436	1.350	1.477	1.465	1.430
Collaborative filtering algorithm for reinforcement learning	1.098	1.108	1.063	1.075	1.111	1.091

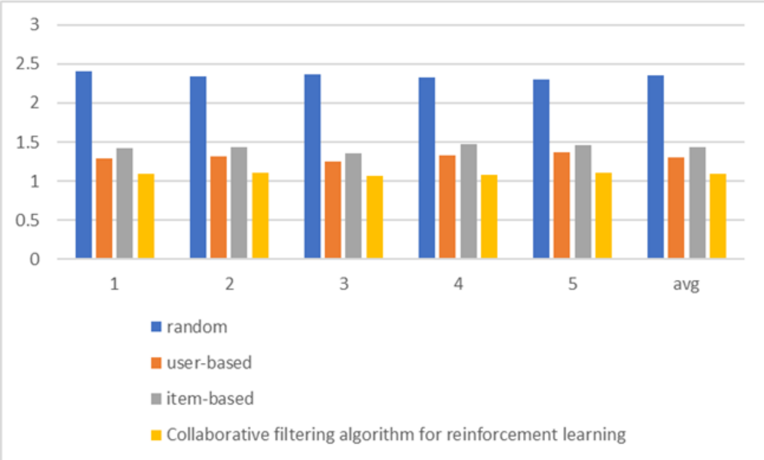

Fig. 5.1: Graphical representation

effect in the field of information retrieval is used as the evaluation standard:

$$\text{errors } = \sqrt{\frac{\sum_{j=1}^{N}[\text{prediction}(j) - \text{real}(j)]^2}{N}}$$

prediction(j) is the expected value of i financial product. real(j) is the objective evaluation of j financial products. N is the estimated number of wealth management products.

5.3. Results and analysis. The three methods are compared and analyzed. Firstly, an evaluation method for calculating random numbers is presented. An evaluation method based on probability distribution is proposed. The method is based on the worst-case baseline. Secondly, user-oriented collaborative filtering (user-based CF) is designed. Then, the resource cooperative filtering method (item-based CF) is proposed. During the experiment, 80% of the video score was scored, and then the collaborative filtering algorithm of reinforcement learning was trained. The top five resources in each user candidate set are predicted using the existing collaborative filtering algorithm of reinforcement learning. And compare it with the actual evaluation value of the user. All nearest neighbor values are 20 when executing both the regular user base CF and item-based CF methods. The information prediction of reinforcement learning collaborative filtering algorithm is presented. Table 5.2 shows the results of the mean deviation test of five different algorithms.

The elements in Table 5.2 represent the mean deviation of the corresponding algorithm in the corresponding data set. Avg is the average score of the five tests. Figure 5.1 gives a graphical representation of Table 5.2. The test results of the five data sets can all show a stable effect. The result shows that the mean value of the random variable is 2.349. A conventional user-based collaborative filtering approach can bring the average down to 1.308. The resource collaborative filtering algorithm can reach 1.430. The collaborative filtering algorithm of reinforcement learning can reduce the mean-variance to 1.091. Two kinds of recommendation accuracy based on collaborative filtering are further improved. The experimental results show that the method is very stable on various samples. In addition, tag information is added to the cooperative filtering. It can improve the user's ability to predict network resources to improve the recommendation accuracy of the network.

6. Conclusion. This paper proposes that the recommendation algorithm of reinforcement learning bank financial products based on user preference and collaborative filtering is feasible. This overcomes the inaccuracy caused by the scarcity of user data. Then, an improved reinforcement learning method is proposed, which can effectively improve the system's performance. Then, use the learned user interest vector and the collaborative filtering method to realize the personalized recommendation to the user. Experimental results show that the enhanced learning method with tags is superior to the conventional collaborative filtering method in recommendation accuracy. Efficient financial product recommendation can also be applied to electronic financial product networks and information searches.

REFERENCES

[1] Thakker, U., Patel, R., & Shah, M. (2021). A comprehensive analysis on movie recommendation system employing collaborative filtering. Multimedia Tools and Applications, 80(19), 28647-28672.
[2] Fkih, F. (2022). Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison. Journal of King Saud University-Computer and Information Sciences, 34(9), 7645-7669.
[3] Bouazza, H., Said, B., & Laallam, F. Z. (2022). A hybrid IoT services recommender system using social IoT. Journal of King Saud University-Computer and Information Sciences, 34(8), 5633-5645.
[4] Sinha, B. B., & Dhanalakshmi, R. (2022). DNN-MF: Deep neural network matrix factorization approach for filtering information in multi-criteria recommender systems. Neural Computing and Applications, 34(13), 10807-10821.
[5] Liao, M., & Sundar, S. S. (2022). When e-commerce personalization systems show and tell: Investigating the relative persuasive appeal of content-based versus collaborative filtering. Journal of Advertising, 51(2), 256-267.
[6] Hanafizadeh, P., Barkhordari Firouzabadi, M., & Vu, K. M. (2021). Insight monetization intermediary platform using recommender systems. Electronic Markets, 31(2), 269-293.
[7] Javed, U., Shaukat, K., Hameed, I. A., Iqbal, F., Alam, T. M., & Luo, S. (2021). A review of content-based and context-based recommendation systems. International Journal of Emerging Technologies in Learning (iJET), 16(3), 274-306.
[8] Lahoud, C., Moussa, S., Obeid, C., Khoury, H. E., & Champin, P. A. (2023). A comparative analysis of different recommender systems for university major and career domain guidance. Education and Information Technologies, 28(7), 8733-8759.
[9] Bellini, P., Palesi, L. A. I., Nesi, P., & Pantaleo, G. (2023). Multi clustering recommendation system for fashion retail. Multimedia Tools and Applications, 82(7), 9989-10016.
[10] Lian, S., & Tang, M. (2022). API recommendation for Mashup creation based on neural graph collaborative filtering. Connection science, 34(1), 124-138.
[11] Karn, A. L., Karna, R. K., Kondamudi, B. R., Bagale, G., Pustokhin, D. A., Pustokhina, I. V., & Sengan, S. (2023). Customer centric hybrid recommendation system for E-Commerce applications by integrating hybrid sentiment analysis. Electronic Commerce Research, 23(1), 279-314.
[12] Zhang, Q., Lu, J., & Zhang, G. (2021). Recommender Systems in E-learning. Journal of Smart Environments and Green Computing, 1(2), 76-89.
[13] Zhu, W. (2021). Topic recommendation system using personalized fuzzy logic interest set. Journal of Intelligent & Fuzzy Systems, 40(2), 2891-2901.
[14] Tahir, S., Hafeez, Y., Abbas, M. A., Nawaz, A., & Hamid, B. (2022). Smart learning objects retrieval for E-Learning with contextual recommendation based on collaborative filtering. Education and Information Technologies, 27(6), 8631-8668.
[15] Harshvardhan, G. M., Gourisaria, M. K., Rautaray, S. S., & Pandey, M. (2022). UBMTR: Unsupervised Boltzmann machine-based time-aware recommendation system. Journal of King Saud University-Computer and Information Sciences, 34(8), 6400-6413.

Edited by: Zhigao Zheng
Special issue on: Graph Powered Big Aerospace Data Processing
Received: Nov 19, 2023
Accepted: Nov 28, 2023