Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 3, pp. 1914–1922, DOI 10.12694/scpe.v25i3.2790

RESEARCH ON DATA MINING AND REINFORCEMENT LEARNING IN RECOMMENDATION SYSTEMS

YUERAN ZHAO*AND HUIYAN ZHAO†

Abstract. This paper aims to help students better grasp the required professional knowledge and core concepts. This paper presents a design method for a multi-layer knowledge base based on XML. According to learners' identity characteristics and learning behaviour, using the mathematical statistics method, the feature expression for the learning system is constructed. Multivariable linear regression theory establishes convergence constraints for accurate and deep mining. The average detection results of the collected samples are used for high-quality deep mining of user portraits in the learning system. This project intends to study the method of solving accurate confidence intervals for user portrait data in the education system. Excel and Access are used to complete the data collection of the teaching object and the construction of the database. A multi-mode interactive editing and processing method of user portrait information for education systems is studied in cloud computing. Finally, a learning system based on mathematical loading mode is proposed, and an object-oriented learning recommendation system is designed. The developed teaching software can enable students to get more teaching guidance when they acquire the required knowledge to improve students' learning effect effectively.

Key words: Knowledge recommendation; Learning needs; Personalization; Learning guidance; Learning system user profile data; Deep excavation; System Design

1. Introduction. Internet autonomous teaching has developed rapidly in the field of computers and the Internet because of its characteristics of "individuality," "autonomy," "initiative," and "non-timeliness." Scholars have built a teaching resource-sharing and management system based on network technology. They set up large-scale open courses for teachers and students so that teachers and students can use the online teaching platforms to carry out interactive learning. Currently, online self-study faces the following problems:

1. learner-oriented online learning environment can not integrate many learning resources well. Students often have difficulties in finding the materials they need for their studies.
2. Students' hidden learning needs cannot be discovered from their behavioral characteristics.
3. Lack of individualized knowledge recommendation and dynamic learning trajectory generation mechanism.
4. There are various forms of research data.

The lack of semantic information makes it difficult for computers to understand and automate. Reference [1] uses OpenCL to implement the KNN parallel computation method. The fine-grained parallelism method and the improvement of multiple thread sets are adopted in the ranging process. In the classification process, the memory model is doubly classified, and the depth of classification is increased. However, this algorithm has a high amount of computation and a substantial real-time. Literature [2] provides an effective NB-MAFIA (Maximum Logistic Organization System) method. Using the compressional coefficient of the N-List and the effective cross-algorithm, the support of the item collection can be solved quickly. Pruning the search space and discovering supersets are used to improve the performance of this method. The performance of this algorithm is not ideal for deep data mining. This paper uses mathematical statistics to build a learning system feature model based on the learner's identity characteristics and learning behavior characteristics [3]. Multivariable linear regression theory establishes convergence constraints for accurate and deep mining. The average detection results of the collected samples are used for high-quality, in-depth mining of the user portraits of the learning system. The behavioral characteristics of learners are intensely studied. Excel and Access are used to complete

*Academic Affairs Office, Zhengzhou Shengda University, Zhengzhou 451191, China (Corresponding author, rachel_yr2023@126.com)
†School of Information Engineering, Zhengzhou Shengda University, Zhengzhou 451191, China

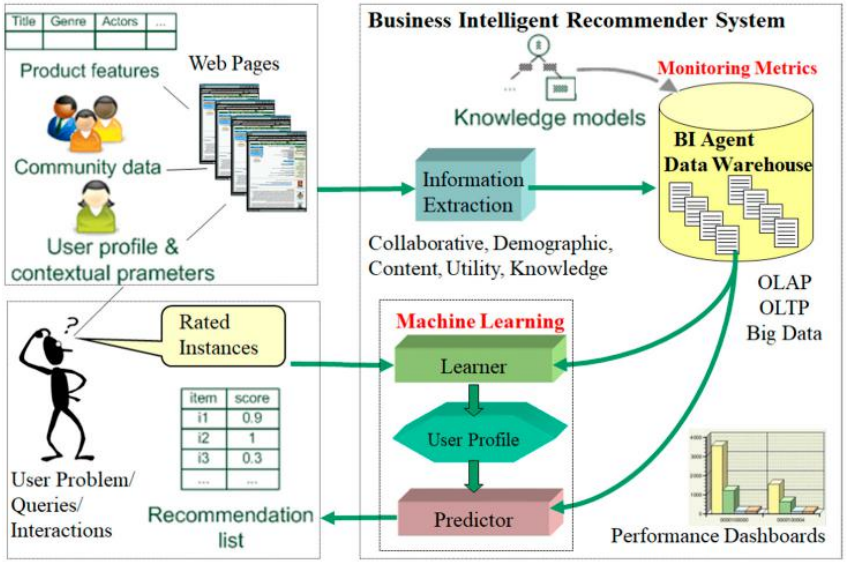

Fig. 2.1: Framework of learning recommendation system based on data mining.

the data collection of the teaching object and the construction of the database. A multi-mode interactive editing and processing method of user portrait information for education systems is studied in cloud computing.

2. Learn the design of the recommendation system. In essence, the learning recommendation system is realized through machine learning. Through the collection of students' basic information and analysis of students' learning habits, learning behaviors and test results, the external and internal learning needs of students are identified [4]. Students actively seek knowledge in the knowledge base to meet their needs—a dynamic generation of learning paths to facilitate learners to complete learning better.

2.1. Overall system framework. The general framework of the learning recommendation system is shown in Figure 2.1 (image cited in Informatics 2017, 4(4), 40). The basic steps of this method are as follows:

1. For the students who adopt this method for the first time, a questionnaire survey is conducted, and their basic information is registered. Through the self-introduction of the students and the analysis of the questionnaire, the students' subject is classified. It includes the relevant body of knowledge, cutting-edge information, and essential citations. In this way, the basic knowledge can be deduced.
2. Students can evaluate the knowledge points presented. All information is stored in the student's database and the student's learning behavior database.
3. Identify students' possible learning interests through users' basic information, learning behavior, and exam results.
4. Make knowledge recommendations based on user feedback, combined with the user's learning requirements, learning performance and the interrelation between knowledge points.
5. Students will learn the next course or topic according to the information the system pushes.

2.2. Building a Knowledge Base. Through the establishment of students' personal information and learning behaviour database, students' learning needs can be found. These include students' personal information, evaluation of learning performance and so on. All the data are stored in the database. The knowledge base includes a central course, knowledge points and related teaching and research resources [5]. Among them, knowledge processing mainly classifies, organizes and transforms the data to form a multi-level knowledge base based on XML. The hierarchy of the knowledge base is shown in Figure 2.2.

Expertise is the highest level in the hierarchy of the knowledge base. The XML file uses majorList.xml to describe the specific content in each discipline. The course information is described in the XML file [6]. Each

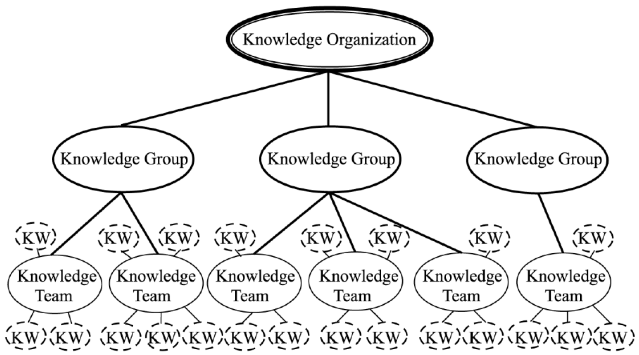

Fig. 2.2: Knowledge base hierarchy diagram.

lesson has a knowledge tree of chapters. XML is represented as separate parts. Knowledge in all sections of the XML file is represented by "partial nodes." Some nodes support hierarchical nested structures. The knowledge points in each chapter include more than one meta-knowledge point. Use a separate knowledge section ID.xml representation. The meta-knowledge point is the smallest unit of knowledge point. It can no longer be divided. Each meta-knowledge node includes related teaching resources, research resources, background resources, expansion resources and so on. This information comes in a variety of ways. These files can be Word, PDF, text, videos, etc. When describing the characteristics of chapter and meta-knowledge, it should consider not only the number, name, keywords, difficulty, and importance but also the interrelation of knowledge.

3. Data deep mining algorithm.

3.1. Mathematical modeling of multiple linear regression. The user portrait of the learning-oriented system is deeply analyzed using the sample mean detection method, and its stability is analyzed [7]. Suppose that the deep mining of the user profile data in the learning system makes the statistical function $g(u), g(u,e)$ continuously bounded on the range of U. It is expressed by $g(U), g(U,e)$. Since $G(U)$ is the unique minimum modular eigenvalue of $g(u)$. Therefore, the initial value of the Gaussian function, which trains the user portrait data for in-depth mining, is expressed as follows:

$$\bar{g}_e^{(0)} = \bar{g}_e = g(n(U^{(0)}), e^{(0)}) \tag{3.1}$$

The principle of mathematical statistics is used to construct a state characteristic equation for the user portrait data in the learning system:

$$\inf G(U,e) - \inf g(U,e) \leq \frac{T}{2}||C(U)||_\infty^2 \tag{3.2}$$

The results show that all the constraints $\ln x, e^u$are monotonically growing functions in the regressive distribution space. In the homogeneous Sobolev space, the order of the user portrait data used in the learning system is intensely mined, and its weight is:

$$C(G(U,e)) - C(g(U,e)) \leq T||C(U)||_\infty^2 \tag{3.3}$$

The robust optimal solution of user portrait data for the learning system is obtained using the sample mean detection method [8]. This project intends to use deep neural networks for research. Deep learning processing of data mining is expressed through the Jacobian matrix $H(u)$:

$$H(u) = \begin{pmatrix} \frac{\partial z_1(u)}{\partial u_1} & \frac{\partial z_1(u)}{\partial u_2} & \cdots & \frac{\partial z_1(u)}{\partial u_n} \\ \frac{\partial z_2(u)}{\partial u_1} & \frac{\partial z_2(u)}{\partial u_2} & \cdots & \frac{\partial z_2(u)}{\partial u_n} \\ \vdots & \cdots & \ddots & \cdots \\ \frac{\partial z_N(u)}{\partial u_1} & \frac{\partial z_N(u)}{\partial u_2} & \cdots & \frac{\partial z_N(u)}{\partial u_n} \end{pmatrix} \tag{3.4}$$

This project intends to study an accurate method for solving user portrait data in the education system. The vector function of edge solution for deep data mining is:

$$c_{ji}(t+1) = c_{ji}(t) - \lambda \frac{\partial G}{\partial c_{ji}} \tag{3.5}$$

$$y_{tj}(t+1) = y_{tj}(t) - \lambda \frac{\partial G}{\partial y_{tj}} \tag{3.6}$$

At the balance point $Q_0(u_1^0, u_2^0)$ of the delay discontinuity, the spectral features of the user portrait in the learning system are extracted. The output spectral characteristic quantity is obtained:

$$\dot{u}(t) = \alpha u(t) + \theta u(t - s_1(t) - s_2(t)) \tag{3.7}$$

where $u(t) = \gamma(t), t \in [-h, 0]$. By using deep learning, the data mining process is adaptive and optimized. This paper obtains the training vector:

$$u(t) = (u_0(t), u_1(t), \cdots, u_{t-1}(t))^T \tag{3.8}$$

3.2. Learning system user portrait data mining output.

Optimization solution of stable features for deep mining of user portrait data of the learning system. The existing SVM algorithm and BP neural network algorithm have significant differences in the recognition accuracy of different samples due to the interference of sampling data. Text recognition based on a deep confidence network can be divided into two stages: pre-training artificial neural network and network adjustment [9]. Most existing classification methods use dimensionality reduction to avoid dimensionality disaster, while deep belief networks (DBN) can extract low-dimensional features with strong discrimination ability from massive original features. In this way, the classification model can be built directly without dimensionality reduction. Meanwhile, it fully uses the rich information in the text. The weights of each BP neural network level are initialized using DBN network weights [10]. This method does not need to initialize any initial value of DBN, nor does it need to extend the BP neural network. BP neural network is used for global optimization to solve the local extreme value problem caused by DBN's randomness of weight parameters.

The robust optimal solution of user portrait data for the learning system is obtained using the sample mean detection method [9]. In the probability distribution interval, the user portrait data of the learning system is intensely mined. Divide the initial value $U^{(0)}$ of the initial cluster center into N parts $U^{(1)}, U^{(2)}, \cdots, U^{(N)}$,$U^{(0)} = \bigcup_{i=1}^{N} U^{(i)}$. This project intends to study an accurate method for solving user portrait data in the education system. A boundary value for convergence can be obtained:

$$C_N = \max_{1 \le i \le N} ||C(U^{(i)})||_\infty \tag{3.9}$$

The first-order inertial output vector of deep learning is $\omega_{tt} - \Delta\omega + |\omega|^p \omega = 0, (p > 4)$. And when$N \to \infty$ satisfies $e_N \to \infty$. An algorithm based on mathematical statistics is proposed and used to optimize the user characteristics in the learning system. Deep learning convergence can be expressed as follows:

$$s_j = \sum_{i=0}^{t-1} (u_i(t) - \eta_{ij}(t))^2, j = 0, 1, \cdots, N-1 \tag{3.10}$$

Where $\eta_j = (\eta_{0j}, \eta_{1j}, \cdots, \eta_{t-1,j})^T$,$\forall \varepsilon > 0, \exists \hat{N} > 0$. When $N > \hat{N}$ is $|\min_{u \in U^{(0)}} e = (g(u) - \lambda_N)| < \varepsilon$. The following formula describes the convergence and optimality of data mining.

$$\min_{0 \le \lambda_i \le c} C = \frac{1}{2} \sum_{i,j=1}^{l} v_i v_j \lambda_i \lambda_j T(u_i, u_j) - \sum_{i=1}^{l} \lambda_i + b \left(\sum_{i=1}^{l} v_j \lambda \right) \tag{3.11}$$

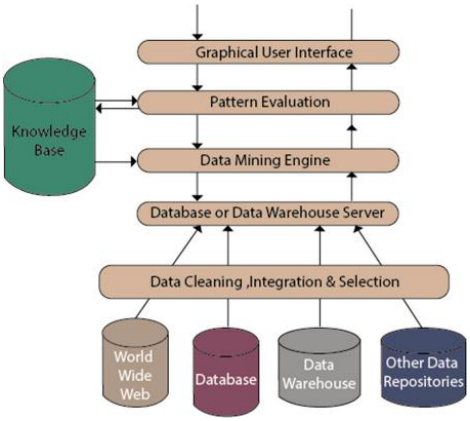

Fig. 3.1: Data mining design architecture flow.

Where (u_i, u_j) represents a controller with linear convexity. It satisfies $c \in (a_1, a_N]$ and satisfies the convergence range:

$$\bar{\delta}_j \leq \frac{f_i(\lambda) - f_i(\beta)}{\lambda - \beta} \leq \delta_j^+ \tag{3.12}$$

The average index $s_c = \frac{s-1}{2}$ is determined for the state feedback controller. The upper bound is taken according to the conditional variance$R_z^i(t)$ of boundary convergence for deep mining of user portrait data in the learning system.

$$\lambda = (\lambda_1, \lambda_2, \cdots, \lambda_n) \neq 0 \tag{3.13}$$

$$\lambda^T L \lambda = \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i \lambda_j L_{ij} \geq 0 \tag{3.14}$$

In the learning system, user portrait data collection point is limited.

1) $\lim\limits_{n\to\infty} \sup |g^n(u) - g^n(v)| > 0, \quad \forall u, v \in R, u \neq v;$
2) $\lim\limits_{n\to\infty} \inf |g^n(u) - g^n(v)| = 0, \quad \forall u, v \in R;$
3) $\lim\limits_{n\to\infty} \sup |g^n(u) - g^n(v)| > 0, \quad \forall u \in R, \forall v \in Q(g)$

The deep confidence interval of user portrait in the education system is solved accurately. At the same time, the behaviour characteristics of learners are deeply explored.

Implementation of data collection flow chart. The process of data mining is designed. The deep mining system of learner portrait data based on embedded processor is studied. The system uses Lab Window/CVI as the software development tool. The deep mining technology of learner portrait data based on embedded Linux is studied. The wireless communication of the network is realized by using ZigBee protocol. The establishment of basic database is based on IEEE802.15.4 technical specification. Adaptive learning method is used in data acquisition to realize the optimal control of data acquisition [10]. Set the sample clock for the A/D module. Adds the required data sources to the user portrait dataset. The software loading function is realized by using the idea of embedded network service. This system can collect and process the learner's portrait data. Based on the above analysis, the design framework for deep mining of user portrait data in the learning system is obtained (Figure 3.1).

Table 4.1: Experimental data set.

Serial number	Data set	Number of transaction items	Number of transaction records
S1	T25I10D10K	1031	5104
S2	Retail	17156	91836
S3	Musroom	124	8463
S4	Kosarak	42990	1031252

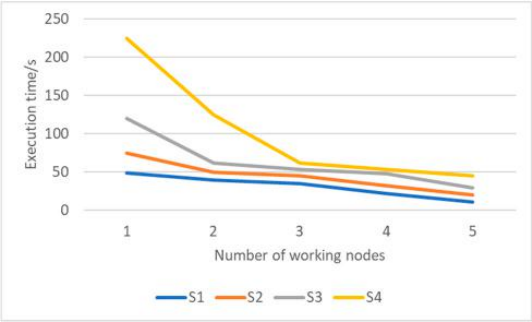

Fig. 4.1: Algorithm execution time.

4. Experiment and analysis. This project adopts the experimental clustering platform built by the Tongfang TR730 series server of Tsinghua University. One controller node and four working nodes are virtualized. Each node has 24 cores and 64 GB of storage [11]. All computing nodes are performed on Ubuntu16.10 OS. JDK1.8 and Eclipse compilation methods are used. Spark2.1.0 and Hadoop2.7.3 construct the cloud computing platform based on Hadoop2.7.3.

4.1. Dataset. The paper will collect and validate four distinctive big data sets on the mechanical data analysis and data mining platforms of FIMI and SPMF. The characteristics of this data collection are shown in Table 4.1.

S1 datasets are collections of artificial data generated by a random transaction database generator. S2 data refers to the retail data used in the supermarket basket mode. It keeps detailed records of customers' business in the mall. S3 is an open fungus data collection. The S4 data set provides a click action for a specific web news entry.

4.2. Experiment and analysis.

Scalability analysis of the algorithm. The experiment evaluated the algorithm's scalability by increasing the number of working nodes and replicating the original data set. The algorithm is dynamically tracked and improved to keep the data quantity constant [12]. Figure 4.1 shows that as the number of nodes changes from 1 to 2, the execution time of the method decreases almost linearly with the increase of the number of nodes. In the process from 3 to 5 nodes, the effect of this method is not significant. This shows that the performance of this method in parallelization has reached a relatively high level.

Under the condition that the system cluster size is five working nodes constant, the corresponding data sets are copied respectively [13]. The speed of the method changes as the data set size increases. Figure 4.2 shows that the running time of this method approximates a straight-line increase as the data set increases. At the same time, this project also proposes a parallel architecture suitable for mass data at various scales.

Algorithm performance analysis. The paper will use this paper's support vector machine, neural network, and nonlinear data mining algorithms to analyse four data types. The experiment was repeated five times. The average execution time of the algorithm is used as the final result to evaluate the performance of the algorithm. Figure 4.3 shows data set S1 with the minimum support set at 0.10%. The results show that

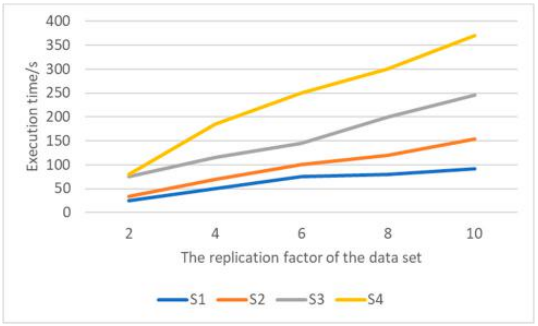

Fig. 4.2: Algorithm execution time.

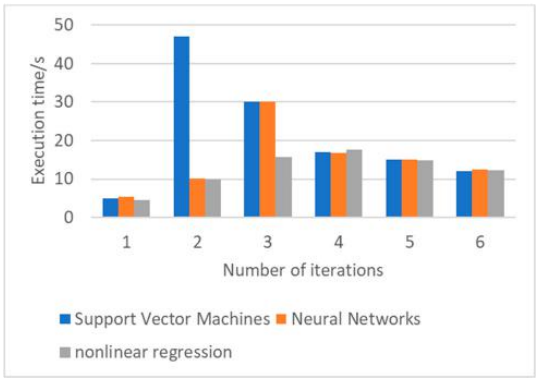

Fig. 4.3: Performance analysis of different algorithms on the S1 dataset.

the second iteration performs better than the SVM algorithm. In the third iteration, the performance of this algorithm is better than that of the two algorithms. After three iterations, the problems to be solved become small and stable [14]. When the number of frequent items in the iteration is small, the algorithm platform adaptively selects the traditional strategy to process the candidate set. In this case, the performance of the three algorithms is roughly the same.

Figure 4.4 shows the data set S2 with a minimum support pre-set of 0.15%. The results show that the proposed method has more advantages than the SVM method in the second iteration. Each iteration's performance is similar to that of the neural network algorithm [15]. This is because the set of pending objects gets smaller and becomes stable after two iterations. The number of occurrences in the last iteration process is small so that the algorithm can choose the standard solution according to the actual situation. This method is similar to the support vector machine and neural network algorithms.

Figure 4.5 shows the data set S3 with the minimum support set to 30% in advance. The results show that the proposed method has more advantages than the traditional neural network method in the second iteration. It is consistent with the iterative results of the support vector machine. This is because the set of pending objects becomes smaller and tends to be stable after two iterations. If the number of items often appearing in the iteration process is minimal, the method can choose the appropriate method to deal with the candidate set according to the need. In this case, the performance of the three algorithms is roughly the same.

Figure 4.6 shows data set S4 with the minimum support set at 0.60 percent. The results show that the nonlinear data mining method is much better than the SVM method in the second iteration. Because of the large scale of D4 samples, the project will select corresponding optimization strategies in 3,4,5 and 6 iterations. In the process of frequent monomial storage and transaction pruning, the execution time of the algorithm is

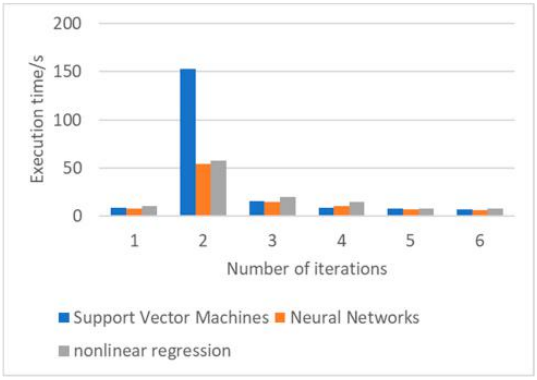

Fig. 4.4: Performance analysis of different algorithms on the S2 data set.

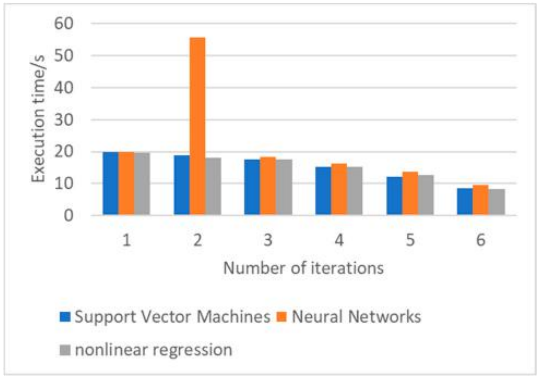

Fig. 4.5: Performance analysis of different algorithms on the S3 dataset.

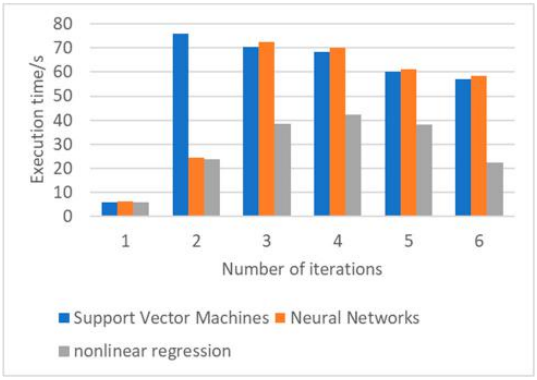

Fig. 4.6: Performance analysis of different algorithms on the S4 dataset.

shortened by using the Split Bloom Filter. This makes the performance of nonlinear data mining algorithms better than support vector machines and neural network algorithms.

5. Conclusion. This paper studies the efficient mining method of education-oriented user portrait data and integrates it with deep learning and adaptive learning to improve the information acquisition ability of

education-oriented user portrait data. The user portrait data in the learning system is analysed in depth. The experiment proves the excellent applicability of this method. It is a scalable distribution algorithm. For large data sets of different sizes, the nonlinear data mining algorithm performs similarly or better than the support vector machine and neural network algorithms on the Spark platform.

REFERENCES

[1] Nitu, P., Coelho, J., & Madiraju, P. Improvising personalized travel recommendation system with recency effects. Big Data Mining and Analytics, 2021; 4(3): 139-154.
[2] Al Fararni, K., Nafis, F., Aghoutane, B., Yahyaouy, A., Riffi, J., & Sabri, A. Hybrid recommender system for tourism based on big data and AI: A conceptual framework. Big Data Mining and Analytics, 2021; 4(1): 47-55.
[3] Javed, U., Shaukat, K., Hameed, I. A., Iqbal, F., Alam, T. M., & Luo, S. A review of content-based and context-based recommendation systems. International Journal of Emerging Technologies in Learning (iJET): 2021; 16(3): 274-306.
[4] Singh, P. K., Pramanik, P. K. D., Dey, A. K., & Choudhury, P. Recommender systems: an overview, research trends, and future directions. International Journal of Business and Systems Research, 2021; 15(1): 14-52.
[5] Steck, H., Baltrunas, L., Elahi, E., Liang, D., Raimond, Y., & Basilico, J. Deep learning for recommender systems: A Netflix case study. AI Magazine, 2021;42(3): 7-18.
[6] Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., & He, X. Bias and debias in recommender system: A survey and future directions. ACM Transactions on Information Systems, 2023; 41(3): 1-39.
[7] Baczkiewicz, A., Kizielewicz, B., Shekhovtsov, A., Watróbski, J., & Sałabun, W. Methodical aspects of MCDM based E-commerce recommender system. Journal of Theoretical and Applied Electronic Commerce Research, 2021;16(6): 2192-2229.
[8] Deldjoo, Y., Noia, T. D., & Merra, F. A. A survey on adversarial recommender systems: from attack/defense strategies to generative adversarial networks. ACM Computing Surveys (CSUR): 2021; 54(2): 1-38.
[9] Joe, M. C. V., & Raj, D. J. S. Location-based orientation context dependent recommender system for users. Journal of Trends in Computer Science and Smart Technology, 2021; 3(1): 14-23.
[10] Wu, D., Shang, M., Luo, X., & Wang, Z. An L 1-and-L 2-norm-oriented latent factor model for recommender systems. IEEE Transactions on Neural Networks and Learning Systems, 2021; 33(10): 5775-5788.
[11] Ferrari Dacrema, M., Boglio, S., Cremonesi, P., & Jannach, D. A troubling analysis of reproducibility and progress in recommender systems research. ACM Transactions on Information Systems (TOIS): 2021; 39(2): 1-49.
[12] Huang, Z., Liu, Y., Zhan, C., Lin, C., Cai, W., & Chen, Y. A novel group recommendation model with two-stage deep learning. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021; 52(9): 5853-5864.
[13] Yue, W., Wang, Z., Zhang, J., & Liu, X. An overview of recommendation techniques and their applications in healthcare. IEEE/CAA Journal of Automatica Sinica, 2021; 8(4): 701-717.
[14] Elbadawi, M., Gaisford, S., & Basit, A. W. Advanced machine-learning techniques in drug discovery. Drug Discovery Today, 2021;26(3): 769-777.
[15] Ageed, Z. S., Zeebaree, S. R., Sadeeq, M. M., Kak, S. F., Yahia, H. S., Mahmood, M. R., & Ibrahim, I. M. Comprehensive survey of big data mining approaches in cloud systems. Qubahan Academic Journal, 2021; 1(2): 29-38.

Edited by: Zhigao Zheng
Special issue on: Graph Powered Big Aerospace Data Processing
Received: Nov 2, 2023
Accepted: Nov 20, 2023