Computer Fraud and Security
ISSN (online): 1873-7056

Research on Library Knowledge Management System Based on Computer Data Information Security

Yunhui Zhang[1,*], Haoyu Li[2]

[1]School of Economics and Management,Suihua University , Suihua, Heilongjiang , China
[2]School of agriculture and water conservancy project ,Suihua University , Suihua, Heilongjiang , China
**Corresponding author*

Abstract:

In order to improve the effect of library knowledge management, this paper combines data mining technology to conduct library knowledge management data mining, and constructs a library knowledge management performance evaluation system. Moreover, this paper proposes a subspace clustering algorithm based on spectral clustering for sparse sample self-expression. In the performance evaluation of library knowledge management, this paper uses subspace clustering to cluster high-dimensional data, and maintains the essential connection of sample points to achieve better clustering effect. In addition, on the basis of applying knowledge management and data mining technology, this paper proposes the process of library knowledge management system based on data mining according to the actual work requirements of the library. Through the experimental research, it can be seen that the performance evaluation method of library knowledge management based on data mining proposed in this paper can effectively improve the efficiency of library management.

Keywords: data mining; library; knowledge management; performance evaluation; Information Security

1 INTRODUCTION

Library management based on knowledge management refers to the process of applying knowledge management theory, technology and methods to improve library management work, enhance library management functions, and achieve library management goals. It is the specific application and practice of knowledge management in library management. The library management based on knowledge management incorporates "knowledge" into the management object and expands the library management object in the traditional sense. That is, the library management object based on knowledge management is composed of document information, personnel, individual and collective knowledge, technology, management system, funds, equipment, architecture, environment and other factors. The main task of library management based on knowledge management is to rationally organize and utilize the existing resources of the library so that they can fully play their role. Moreover, it promotes the exchange, sharing and utilization of knowledge (including explicit knowledge and tacit knowledge) of employees within the library, and expands the knowledge capital of the library. Further, it builds a learning-oriented organizational structure, optimizes library business processes, improves work efficiency, and strengthens library human resource management to improve staff quality and ability. At the same time, it creates an innovative library culture, shapes a knowledge management-oriented library management culture, expands knowledge services, and improves service levels and levels. In addition, it evaluates library management practices, improves management level and improves management efficiency.

The role of library management based on knowledge management is to support the library to achieve its strategic goals, and to assist the library to improve its ability to achieve its strategic goals. Knowledge management is a complex system engineering, and its introduction and implementation are also a step-by-step process. However, as a new management activity, it also has its own specific goals or direct goals, so it is very important to determine the goals of knowledge management. The reason is that the goal is the starting point of all management activities, the basis of PDCA management cycle, and the standard of management performance evaluation. Moreover, if there is no goal system, performance evaluation will be impossible.

From the perspective of library functions, "user-centered" is one of the highest concepts of modern libraries, and the survival and development of libraries must be based on user satisfaction. Therefore, clarifying the knowledge needs of users is a prerequisite for improving library knowledge services. The ways of knowledge service can be summarized as four ways: knowledge navigation, knowledge consultation, knowledge integration and knowledge marketing. When users are faced with how to capture and extract the required information

content from the vast and complex information ocean, the knowledge navigation service reorganizes or innovates the information to generate corresponding knowledge or solutions. Knowledge consultation is based on knowledge, relying on the knowledge, experience and skills of experts and with the help of certain means, to analyze and research the problems, topics or projects raised by users, and to propose suggestions, solutions and measures to solve the problem. The library uses modern information technology to integrate knowledge navigation, knowledge consultation and other services to provide users with knowledge integration services. Knowledge marketing means that the library provides users with the investigation and analysis of the knowledge product market, and participates in the formulation of product prices and the establishment of distribution channels and other related affairs. Moreover, it creates market demand through product knowledge propaganda, realizes the commercialization and market value of knowledge products, and improves the efficiency of the library.

This paper combines data mining technology to conduct library knowledge management data mining, and constructs a library knowledge management performance evaluation system to improve the efficiency of library knowledge management.

2 RELATED WORK

Literature [1] believes: "Knowledge management is a kind of catering measure for the important aspects of the organization's adaptability, organization's survival and competitiveness in the face of increasingly discontinuous environmental changes. In essence, it It includes the development process of the organization, and seeks to organically combine the processing capabilities of data and information provided by information technology and the ability of people to invent and create. Literature [2] believes that knowledge management is the use of collective intelligence to improve strain and innovation ability]. Literature [3] believes: "Knowledge management is the systematic management of important knowledge and related processes such as creation, collection, organization, dissemination, utilization and publicity. It requires the transformation of individual knowledge into group knowledge that can be widely shared and appropriately utilized by an organization. The American Productivity Center defines it as: "Knowledge management is a conscious organizational strategy that ensures that the right knowledge is delivered to the right people at the right time, facilitates its sharing, and puts it into practice in different ways. , and ultimately achieve the purpose of improving organizational performance [4]. Literature [5] believes: "Knowledge management is the effective management of knowledge by using advanced technology in order to promote the creation, access and reuse of knowledge. The literature [6] pointed out that knowledge management in a narrow sense mainly refers to the management of knowledge itself, including the management of knowledge creation, acquisition, processing, storage, dissemination and application. Knowledge management in a broad sense includes not only the management of knowledge, but also the management of knowledge. It also includes the management of various resources and intangible assets related to knowledge, involving the all-round and whole-process management of knowledge organization, knowledge facilities, knowledge assets, knowledge activities, and knowledge personnel. Literature [7] believes: "Knowledge management is a The unification of the management of information and technology, namely knowledge and the management of people, requires the connection of information and information, information and activities, and information and people to form a knowledge network, realize knowledge sharing, and change the traditional organizational structure. Adapt to the emergence and development of 'knowledge workers'. Literature [8] believes that knowledge management is a new management model, and knowledge management is a part of modern enterprise management, that is, setting up a knowledge manager, using modern information technology, and training all aspects of the internal knowledge of the enterprise and the training of employees. management, so that all kinds of knowledge can be effectively used and transformed into greater productivity. It can be seen that knowledge management can also be regarded as a new management idea or management theory.

Literature [9] believes that knowledge management is to solve the problems of innovation, organization and diffusion of a large amount of knowledge or a large amount of information/information. way. Literature [10] believes that knowledge management includes two meanings. On the one hand, it refers to the management of information, realizing knowledge-based value-added, discovery, organization, sharing and use of information through intelligent information 'deep processing; management to realize the discovery, organization, sharing and utilization of tacit knowledge.

Literature [11] believes that library knowledge management refers to the application of knowledge management theories, technologies and methods by libraries to rationally allocate and use knowledge and related resources, fully meet the changing information and knowledge needs of users, and improve the various aspects of modern libraries. functions and processes to better perform them. In a broad sense, library knowledge management is the management and research of all activities and laws related to knowledge production, acquisition, organization, storage, communication, dissemination and application in the library. It also includes the management of library knowledge capital; library knowledge management in a narrow sense is the management of library knowledge itself, including the management process of knowledge production, acquisition, organization, storage, communication, dissemination and application [12]. Literature [13] believes that the knowledge management of the library should include two levels: First, starting from the social function of the library, in the social knowledge management environment, it should give full play to the knowledge itself, especially the sorting of explicit knowledge. The role of development and management of processing, transmission and utilization, to promote knowledge sharing and utilization, and to promote the organic combination of those who need knowledge and the knowledge they need; second, from the perspective of the library as a social organization, knowledge management should be introduced in its internal management. In this way, the potential ability of librarians can be developed by using a large amount of knowledge, and the overall management level of the library can be improved through the sharing of knowledge. Literature [14] believes that library knowledge management is actually the process of collecting, sorting, storing and using its explicit knowledge and tacit knowledge, and making it fully functional.

3 THE MINING METHOD OF LIBRARY KNOWLEDGE MANAGEMENT PERFORMANCE

The task of subspace clustering is to cluster the entire data set and other unknown data samples.

We assume that $\left\{x_i \in R^D\right\}_{i=1}^N$ is a set of given sample points, and these sample points come from n unknown linear or affine subspaces $\left\{S_j\right\}_{j=1}^n$, respectively. Among them, the dimension d_k of each subspace is also unknown, $d_k = \dim\left(S_k\right)\left(0 < d_k < D, k = 1,2,\ldots,n\right)$. Then, the subspace can be expressed in the following form:

$$S_k = \left\{x \in R^D : x = \mu_k + U_k y\right\}, k = 1,2,\ldots,n \quad (1)$$

In order to better understand the meaning of subspace clustering, this paper uses Figure 1 to explain. In the figure, S_1、S_2 and S_3 respectively represent three subspaces from a given three-dimensional data, S_1 and S_2 are two straight lines, and S_3 is a plane.

We find that sample points in 3D data are 1D and 2D in nature. If we cluster by the distance similarity measurement method according to the traditional clustering, it will obviously destroy the correlation between the sample points (straight line S_1、S_2, plane S_3). The goal of subspace clustering is to make full use of the properties of the samples themselves, and divide the samples in different subspaces into low-dimensional subspaces that are consistent with themselves.

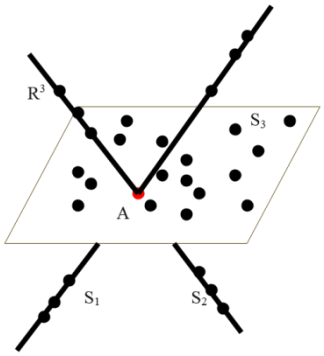

Figure 1 A set of 3D sampling points from three different subspaces (S_1 and S_2 are lines, S_3 is a plane)

Through the definition of subspace, it can be found that when there is n=1, the goal of subspace clustering is to find vector $\mu \in R^{\mathrm{D}}$, subspace benchmark $U \in R^{D \times d_k}$, low-dimensional representation $Y = [y_1, y_2, \ldots, y_N] \in R^{d \times N}$ and subspace dimension d.

Accuracy refers to how accurately the predicted results match the true results, and can be defined as follows:

$$\mathrm{precision}(PC_s) = n(PC_s \cap TC_s) / n(TC_s) \quad (2)$$

The recall rate refers to the proportion of the number of correctly predicted s classes to the total number of s classes in the prediction results, which can be defined as:

$$\mathrm{recall}(PC_s) = n(PC_s \cap TC_s) / n(PC_s) \quad (3)$$

Moreover, the F value is an evaluation standard for comprehensively measuring the recall rate and accuracy rate, and its definition is as follows:

$$F(PC_s) = \frac{2 \cdot \mathrm{precision}(PC_s) \cdot \mathrm{recall}(PC_s)}{\mathrm{precision}(PC_s) + \mathrm{recall}(PC_s)} \quad (4)$$

The above analysis is the indicator of a single s class. If the entire data class needs to be evaluated, we can evaluate it by taking the average of all classes, which is in the following form:

$$\bar{P} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{precision}(PC_s) \quad (5)$$

$$\bar{R} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{recall}(PC_s) \quad (6)$$

$$\bar{F} = \frac{1}{n} \sum_{i=1}^{n} F(PC_s) \quad (7)$$

In this paper, the clustering error (CE) is used as the evaluation index of the clustering effect. The clustering error is the minimum error between the matching prediction result and the real result under the optimal permutation, and its definition is as follows:

$$CE = 1 - \frac{1}{N} \sum_{i=1}^{N} \delta(tc_i, \mathrm{map}(pc_i)) \quad (8)$$

We assume that $X_i \in R^{D \times N_i}$ represents N_i points in subspace i, and d_i is the rank of matrix X_i. The

columns of the sample matrix are arranged according to the matrix set $X\Gamma = [X_1, X_2, \ldots, X_n]$ of n subspaces, among them, $\Gamma \in R^{N\times N}$ is a matrix of unknown arrangement. Then, the matrix X_i can be decomposed into the following form:

$$X_i = U_i Y_i i = 1, 2, \ldots, n \quad (9)$$

Among them, $U_i \in R^{D\times d_i}$ represents an orthonormal basis of subspace i, and $Y_i \in R^{d_i \times N_i}$ is a low-dimensional representation of the sample points in U_i. If the subspaces are independent, the rank $r = \text{rank}(X) = \sum_{i=1}^{n} d_i \le \min\{D, N\}$ of the matrix X can be expressed as:

$$X\,\Gamma = [U_1, U_2, \ldots, U_n] = \begin{bmatrix} Y_1 & & & \\ & Y_2 & & \\ & & \ddots & \\ & & & Y_n \end{bmatrix} = UY \quad (10)$$

Among them, there is $U \in R^{D\times r}, Y \in R^{r\times N}$.

Generalized Principal Component Analysis (GPCA) is a well-known polynomial-based algebraic method. Its main principle is to use a force-degree polynomial,and to derive the normal vector containing this point at a given point. Thus, subspace clustering can be achieved by grouping these normal vectors using certain techniques.

The specific implementation process of GPCA is mainly divided into the following three steps: In the first step, GPCA projects the data samples into an $r = d_{\max} + 1$-dimensional R^D subspace. The second step is to find a homogeneous polynomial of degree n to fit the projection matrix. The specific equation can also be transformed into the following quadratic curve form:

$$c_1 x_1^2 + c_2 x_1 x_2 + c_3 x_1 x_3 + c_4 x_2^2 + c_5 x_2 x_3 + c_6 x_3^2 = 0 \quad (11)$$

Each polynomial is of degree n in x and denoted as $c^{\mathrm{T}} v_n(x)$. Among them, c is a vector of coefficients and $v_n(x)$ is a vector of monomials of degree n for all x. The third step is to calculate the normal vector b_i according to the coefficient vector c.

In statistical learning, individual subspaces are well modeled by Gaussian distributions. It directly assumes that each probability distribution is Gaussian, then a mixed Gaussian model can be used for multiple subspaces, so the subspace clustering problem is transformed into a model estimation problem. However, the expectation-maximization EM method can handle the maximum likelihood estimation of mixed models well, and EM is an iterative process of updating data clustering and model estimation. The Gaussian mixture model is assumed to be:

$$p(x) = \sum_{i=1}^{n} \pi_i G\left(x; u_i, U_i U_i^T + \sigma_i^2 I\right), \quad \sum_{i=1}^{n} \pi_i = 1 \quad (12)$$

Among them, u is the mean and $U_i U_i^T + \sigma_i^2 I$ is the covariance matrix. In the EM method, the probability that sample x_j belongs to a certain subspace i can be estimated as:

$$p_{ij} = \frac{\pi_i G\left(x_j; u_i, U_i U_i^T + \sigma_i^2 I\right)}{p\left(x_j\right)} \quad (13)$$

When the probability p_{ij} of each sample x_j belonging to each subspace i is obtained, we can perform subspace clustering on it according to the traditional algorithm.

It is different from algebra-based and statistics-based subspace clustering algorithms. Algorithms based on spectral clustering are the most widely used algorithms to solve high-dimensional data clustering. Then, for each sample in the data set, a linear combination that belongs to the same subspace is searched. Meanwhile, a similarity matrix is established according to the linear combination, and finally the similarity matrix is processed by spectral clustering to obtain the final clustering result. Figure 2 below is the basic framework of our spectral clustering-based subspace algorithm.

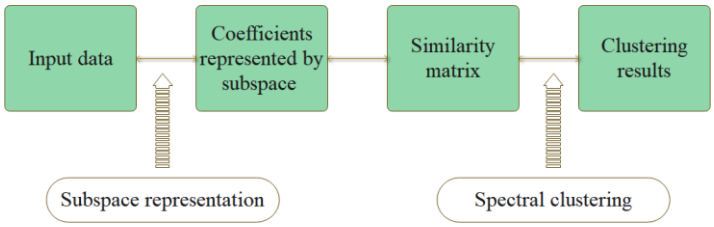

Figure 2 The basic framework of the subspace algorithm based on spectral clustering

The SSC algorithm is to find a sparse linear representation for each sample in the data set, and then use the obtained sparse linear representation coefficients to construct the similarity matrix directly. Its objective function is:

$$\min_{\mathbf{Z}} \| \mathbf{Z}\|_1 + \frac{\lambda}{2} \| \mathbf{X} - \mathbf{XZ}\|_F^2, \text{ s.t. } \operatorname{diag}(\mathbf{Z}) = 0 \quad (14)$$

Among them, Z is the subspace coefficient matrix. The method firstly finds the respective sparse linear representation of each data vector by constructing an l_1-norm of the sparse representation, and the samples in the sparse linear representation all come from the same subspace. Then, the obtained linear representation coefficient matrix is used to construct the similarity matrix as follows:

$$\mathbf{W} = \left(|\mathbf{Z}| + \left|\mathbf{Z}^{\mathrm{T}}\right|\right) / 2 \quad (15)$$

Finally, the matrix W is used to obtain the final clustering result by spectral clustering algorithm. The process of LLR algorithm to achieve clustering is similar to that of SSC algorithm. The main difference is that the LLR algorithm is to find a low-rank representation instead of a sparse representation, that is, to find the matrix with the lowest rank among the matrices composed of all data vectors. However, since the rank minimization problem is an NP-hard problem, LRR replaces $\| Z\|_1$ by the kernel function $\| Z\|_* = \sum \sigma_i(Z)$, among them, $\sigma_i(Z)$ is the i-th singular value of Z. Finally, the objective function can be expressed as:

$$\min_{\mathbf{Z}} \| \mathbf{Z}\|_* + \frac{\lambda}{2} \| \mathbf{X} - \mathbf{XZ}\|_F^2 \quad (16)$$

After the low-rank linear representation matrix Z is obtained, similar to the SSC algorithm, the similarity matrix is established by formula (15) and applied to spectral clustering to complete the final clustering operation.

4 RESEARCH ON THE EVALUATION METHOD OF LIBRARY KNOWLEDGE MANAGEMENT PERFORMANCE BASED ON DATA MINING

4.1 System model building

The library knowledge management system model collects a wide range of knowledge resources in various ways, including a large number of disordered and irrelevant information. Moreover, it generally includes several modules of business process, office automation management, customer relationship management, human resource management, learning and innovation. In addition, it realizes the mutual conversion of tacit knowledge and explicit knowledge in the library. Figure 3 shows the library knowledge management system model based on data mining.

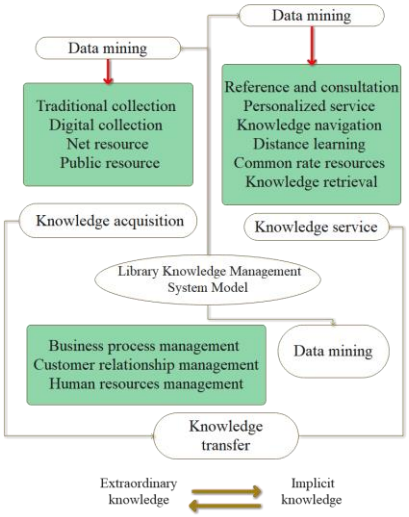

Figure 3 Library knowledge management system based on data mining

Based on the application of knowledge management and data mining technology, and according to the actual work requirements of the library, this paper proposes the process of the library knowledge management system based on data mining, as shown in Figure 4.

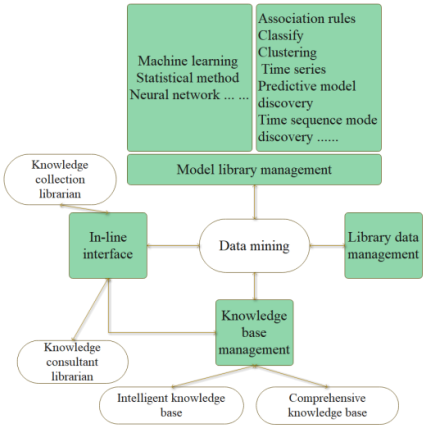

Figure 4 The flow of library knowledge management system based on data mining

The core content of library knowledge management is a management process of knowledge sharing, knowledge operation and knowledge innovation. Moreover, it is the process of organically combining tacit knowledge and explicit knowledge in the library, transforming and interacting with each other, and it is a prerequisite for data mining. The main function of data mining is to provide information reference sources for various scientific research fields. Figure 5 shows the reference consultation process based on data mining.

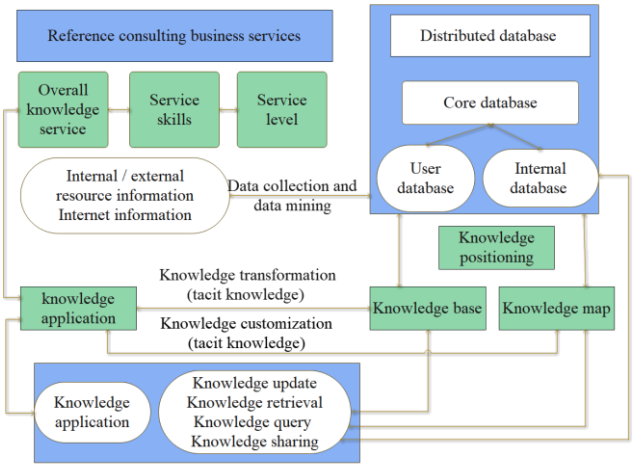

Figure 5 Reference consultation process based on data mining

The knowledge map uses visualization techniques and methods to display various relationships between knowledge and knowledge, between knowledge and people, and between people. The conceptual model of knowledge map is shown in Figure 6.

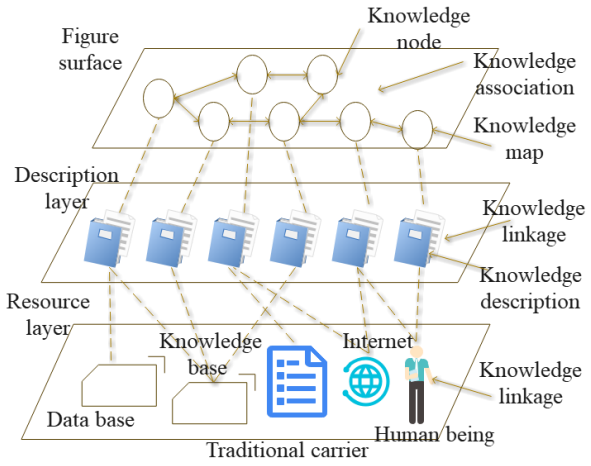

Figure 6 Conceptual model of knowledge map

The design of the prototype system should fully consider the relationship between the whole and the part, the content and the form, so as to integrate the project process type, concept type and professional title type knowledge map organically. The upper layer of the knowledge atlas is connected to the knowledge application components, and the bottom layer is connected to the knowledge base and domain experts, as shown in Figure 7.

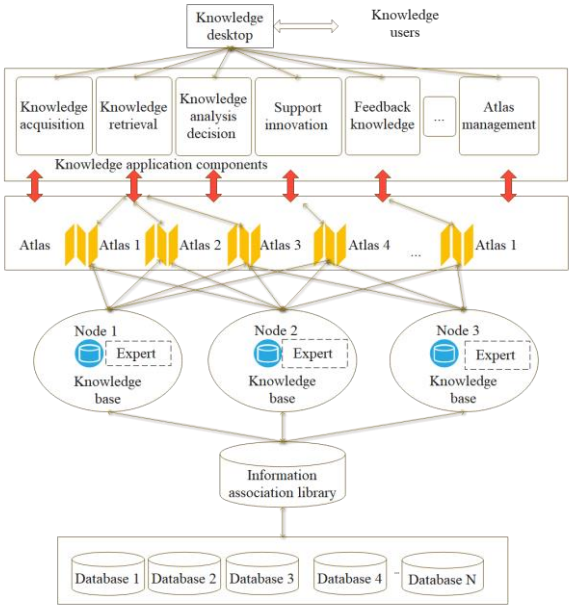

Figure 7 Framework diagram of the prototype system

4.2 System simulation test

This paper conducts performance simulation of the performance evaluation model proposed in this paper, and obtains the simulation diagram as shown in Figure 8.

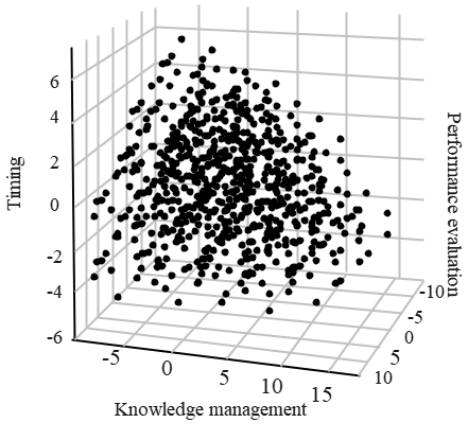

Figure 8 Simulation diagram of performance evaluation model

On the basis of clustering simulation, this paper conducts simulation experiments to verify the performance evaluation method of library knowledge management based on data mining, and counts the data mining effect and performance evaluation effect of this system. The results are shown in Tables 1 and 2.

Table 1 Data mining effect of library knowledge management performance evaluation system based on data mining

Num	Knowledge mining	Num	Knowledge mining
1	92.30	13	89.13
2	89.48	14	90.62
3	91.69	15	90.47
4	87.96	16	90.59

5	92.68	17	88.88
6	86.09	18	85.50
7	90.14	19	87.19
8	86.01	20	92.09
9	91.68	21	91.21
10	87.60	22	92.55
11	86.15	23	90.82
12	92.93	24	88.67

Table 2 Performance evaluation effect of library knowledge management performance evaluation system based on data mining

Num	Performance management	Num	Performance management
1	85.35	13	79.99
2	84.88	14	83.15
3	72.99	15	77.57
4	85.03	16	81.06
5	79.70	17	81.39
6	85.96	18	82.56
7	76.27	19	73.94
8	79.82	20	79.43
9	80.83	21	82.75
10	82.37	22	74.66
11	72.71	23	72.55
12	73.86	24	81.31

Through the above research, we can see that the library knowledge management performance evaluation system based on data mining proposed in this paper can effectively improve the efficiency of library management.

5 CONCLUSION

The goal of library management based on knowledge management is to rationally organize and utilize various library resources, build a learning organization structure, optimize library business processes, and maximize library system functions. Moreover, it helps employees obtain available knowledge to accomplish their tasks more effectively, provides an effective knowledge sharing platform for employees, maximizes the capture, mining, utilization, and dissemination of knowledge, and improves employees' knowledge innovation and knowledge service capabilities. In addition, it better meets the needs of society and users for information and knowledge of the library, and promotes the harmonious development between society and the library, and between users and library staff. This paper combines data mining technology to conduct library knowledge management data mining, and constructs a library knowledge management performance evaluation system. Through the experimental research, it can be seen that the library knowledge management performance evaluation system based on data mining proposed in this paper can effectively improve the efficiency of library management.

ACKNOWLEDGE:

Art Science Planning Project of Heilongjiang Province(Project number:2019D037);Key Topics of Education Science Planning in Heilongjiang Province(Project number:GJB1422383)

REFERENCES

[1] Liu, Y. (2018). Data mining of university library management based on improved collaborative filtering association rules algorithm. Wireless Personal Communications, 102(4), 3781-3790.

[2] Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., Lopez Garcia, A., Heredia, I., ... & Hluchý, L. (2019). Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. Artificial Intelligence Review, 52(1), 77-124.

[3] Mohd Selamat, S. A., Prakoonwit, S., & Khan, W. (2020). A review of data mining in knowledge management: applications/findings for transportation of small and medium enterprises. SN Applied Sciences, 2(5), 1-15.

[4] Papadopoulou, M. D., Kolotourou, K., & Bottis, M. (2021). The Exception of Text and Data Mining from the Academic Libraries Standpoint. Open Journal of Social Sciences, 9(5), 502-539.

[5] Al-Daihani, S. M., & Abrahams, A. (2018). Analysis of academic libraries' facebook posts: text and data analytics. The Journal of Academic Librarianship, 44(2), 216-225.

[6] Yang, J., Li, Y., Liu, Q., Li, L., Feng, A., Wang, T., ... & Lyu, J. (2020). Brief introduction of medical database and data mining technology in big data era. Journal of Evidence-Based Medicine, 13(1), 57-69.

[7] Abu Saa, A., Al-Emran, M., & Shaalan, K. (2019). Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques. Technology, Knowledge and Learning, 24(4), 567-598.

[8] Chaithanapat, P., & Rakthin, S. (2021). Customer knowledge management in SMEs: Review and research agenda. Knowledge and Process Management, 28(1), 71-89.

[9] Saura, J. R. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. Journal of Innovation & Knowledge, 6(2), 92-102.

[10] Cominola, A., Nguyen, K., Giuliani, M., Stewart, R. A., Maier, H. R., & Castelletti, A. (2019). Data mining to uncover heterogeneous water use behaviors from smart meter data. Water Resources Research, 55(11), 9315-9333.

[11] Saura, J. R. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. Journal of Innovation & Knowledge, 6(2), 92-102.

[12] Cominola, A., Nguyen, K., Giuliani, M., Stewart, R. A., Maier, H. R., & Castelletti, A. (2019). Data mining to uncover heterogeneous water use behaviors from smart meter data. Water Resources Research, 55(11), 9315-9333.

[13] Ahmed, B. S., Maâti, M. L. B., & Al-Sarem, M. (2020). Predictive Data Mining Model for Electronic Customer Relationship Management Intelligence. International Journal of Business Intelligence Research (IJBIR), 11(2), 1-10.

[14] Arslan, O., Xing, W., Inan, F. A., & Du, H. (2022). Understanding topic duration in Twitter learning communities using data mining. Journal of Computer Assisted Learning, 38(2), 513-525.