



THE MD-BK-MEANS CONSTRUCTION METHOD FOR LIBRARY READER PORTRAITS

ZHIYU ZHU*

Abstract. Due to the rapid development of internet technology, knowledge acquisition has become more convenient and efficient in network operations. University libraries serve as important resources for readers to acquire knowledge, and online resources and services in libraries have become the main direction for readers to acquire knowledge at present. Research the use of binary K-means clustering algorithm and library reader portrait technology to optimize the design of the reader portrait module and construct a multidimensional and multi perspective reader feature system. Reuse Spark programming language and support vector machine to perform computational processing on reader profile data to ensure accurate segmentation of the dataset. Finally, three datasets were used to test the accuracy and efficiency of the algorithm. The experimental comparison shows that the mining and precision segmentation of parallel SVM on the dataset are 93.20%, 85.16%, and 79.35% on the sample set, respectively, in order to optimize the mining performance of the data. The MD multi view binary K-means algorithm has a total Mahalanobis distance of 3.543, 5.268, and 22.385 on the sample dataset, respectively, to demonstrate its superiority in clustering performance. Therefore, the multi view binary K-means algorithm based on Mahalanobis distance has high advantages in reader portrait technology design, and provides technical support and theoretical reference for library reader portrait technology.

Key words: Portrait technology; Reader characteristic system; K-means clustering algorithm; Multi perspective clustering; Mahalanobis distance

1. Introduction. Library collection resources and book information have always been representative sources of high-quality books. With the development of internet technology, the demand for readers to enter libraries is gradually decreasing, and the resources and services of libraries are also difficult to meet the diverse needs of readers. Therefore, in order to improve personalized services in libraries, in-depth research is conducted on the needs and preferences of readers for book resources to construct a reader behavior database, and to provide readers with high requirements for personalized services. However, the customization of these personalized services did not conduct in-depth analysis of reader behavior and book preferences, thus failing to meet the resource needs and specific services of readers in a timely manner, leading to the waste of book resources. The issue of library education and learning education is highly valued both domestically and internationally, and the utilization of library electronic resources requires in-depth investigation of readers in order to attract them. However, most university libraries lack the utilization of professional book resources for specialized and customized services, and the system service functions of library management personnel also lack scientificity. There are also service problems such as the inability to achieve real-time transformation of the ever-changing needs of readers, insufficient timely updates of book resources, etc. These require in-depth analysis of reader needs and book resource management, and the establishment of a reader behavior database to meet the ever-changing requirements of readers, Furthermore, we provide professional and personalized reader customization services. Based on this, in-depth analysis of the reader profiling technology framework is conducted to obtain reader behavior information, thereby understanding the diverse reading needs of readers. Using optimized binary K-means clustering algorithm to design the portrait module of reader portrait technology, to construct multi perspective and multi-dimensional reader feature information, and finally construct the reader portrait technology system of Backbone Initial K-means clustering algorithm (BK-means) based on Mahalanobis distance, aiming to improve the accuracy of the algorithm and provide technical reference for reader information feature acquisition.

The study consists of four parts. The first part is an elaboration and summary of the current LRPT. The second part is an explanation of LRPT and the construction of a reader behavior feature system, and

*Library, Communication University of Zhejiang, Hangzhou, 310018, China (zzy_cm72@163.com)

the modular design of the reader user system using multi perspective clustering. The third part combines the calculation of the sum of MD to construct a reader portrait technology system and tests and compares it with readers from multiple perspectives, to prove the superiority of its algorithm. The final part is a summary and narration of the entire study.

2. Related works. At present, the LRPT relies on the acquisition and analysis of reader behavior characteristics and algorithm citation, so it requires theoretical and technical support. In recent years, many scholars have conducted extensive research on LRPT and system construction.

Byrkovych et al. proposed to transform the modernization concept of the library industry by combining the current education situation and policies in response to the issue of library and information reform in Ukrainian universities, to assist in the development of the library industry [1].

Wangdi et al. [2] and Van Melik [3] analyzed the necessity of resources for the development of libraries in Bhutan and public libraries in the Netherlands.

Zhou proposes to use blockchain technology to collect user profile information in mobile libraries to promote real-time sharing of profile data [5].

Navarrete et al. proposed editing the main body of the article, recording views and frequencies, and determining museum content to increase online attraction for museums in the English version of Wikipedia [6].

Hamilton et al. proposed using 360 tools to create library service guidance methods and provide technical references for library services [7].

Zeng et al. used K-means to extract image features and construct a model to provide technical support for the protection of Dunhuang culture and heritage in the Smart Library. The object resources of the library are relatively rich, and relevant technologies need to be used to support the development of the library industry [8].

Chen et al. proposed the use of crawler technology and questionnaire surveys to explore the key and difficult points of the construction of smart libraries in universities, thereby providing a theoretical basis for the information construction of university libraries [9].

For the problem of college teaching intelligent service system, Qu uses Big data to analyze mobile learning intelligent services and add resources, teaching models and infrastructure to help personalized services [10].

Ahmad proposed to investigate the preservation of digital resources in university libraries and explore their significant differences, in order to provide protection for the preservation of digital resources [11].

Shiri et al. used digital media to preserve archives and obtain digital interface features to protect cultural heritage. The utilization of library resources by new technologies can also protect library collection resources [12].

Zhang et al. proposed combining time series segmentation and clustering techniques to provide research data for the profiling of attack behavior regarding the issue of path planning intelligent attacks [13].

Li utilizes IoT technology to collect and analyze data from students and faculty for smart campus management systems, in order to improve teaching plans and campus management [14].

Candela G et al. proposed using datasets to create machine operable sets for digital information resources in libraries [15].

As for personalized autonomous learning system, Wang et al. proposed Differential evolution to construct learning path and design system to provide learning resources [16].

Tong et al. designed campus scene logic using virtual logic technology combined with artificial intelligence algorithms to provide reference requirements for campus scene management [17].

Gao proposed a Big data analysis and construction model for short video recommendation strategies to provide theoretical reference for personalized recommendation services [18].

Long uses user-defined input simulations to construct an augmented reality framework to enhance students' interest in learning, thereby increasing learning attractiveness. It can be seen that in addition to user profiling, new technologies can also play a recommendation and personalized service role in other fields [19].

In summary, although many scholars have established many systematic methods for LRPT and its education, there is still a lack of in-depth research on the application of portrait technology and educational resources. Therefore, the use of MD's BK-means algorithm (MD-BK) in this study has innovative advantages in user profiling technology.

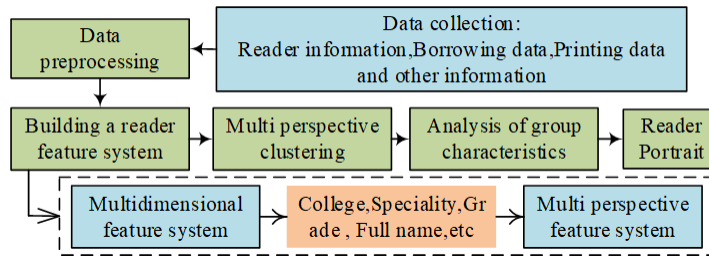


Fig. 3.1: Technology roadmap of Reader Portrait

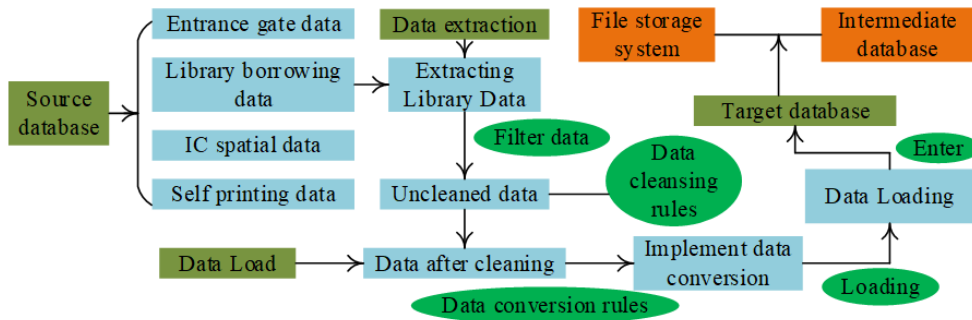


Fig. 3.2: ETL flowchart

3. Model Construction of BK-means Algorithm Based on MD. In order to improve the personalized service quality of library readers, libraries are no longer just the only way to obtain book resources. Due to its lack of precise control over reader needs, data mining techniques and reader profiling techniques can be used to analyze reader behavior. Furthermore, this chapter proposes a multi perspective binary K-means algorithm for MD (MD-MBK-means) to construct a reader feature system.

3.1. LRPT and its feature system construction. University library is an important Big data book resource sharing center, which is rich in resources and high in value, but the diversity of reader needs makes the library service inefficient. Therefore, data mining techniques can be used to collect and process reader behavior information, construct and train models, and then construct reader profiling technology processes to gain a deeper understanding of reader needs. As shown in Fig. 3.1.

In Fig. 3.1, the reader portrait technology roadmap first goes through data processing, builds a feature system through multi-dimensional reader feature information, and then carries out multi perspective clustering and group feature analysis to get the reader portrait. Then, based on the constructed database, the data is effectively extracted to enter the target database, and then extracted through the Extraction-Transformation-Loading (ETL) system. The process is Fig. 3.2.

In Fig. 3.2, the ETL process includes data extraction, cleaning, transformation, and loading. Data cleansing and transformation are carried out according to relevant rules. The transformed data corresponding fields are transformed into the target database to facilitate data preprocessing. The subsequent construction of a multi perspective reader feature system involves dividing readers through information labeling. The reader feature system is divided into linear features (dimensional features such as college, grade, major, and name) and implicit features. It includes the perspective characteristics of reader activity, reader borrowing rate, electronic resource utilization rate, public resource utilization rate, and text features of books borrowed by readers. Finally, the

implicit feature formula is calculated to obtain the Reader Activity (AR) formula (3.1).

$$AR = \frac{T}{D} \quad (3.1)$$

In equation (3.1), T represents the number of times readers have entered the library during the statistical time, and D is the number of valid days for readers in the library. It intuitively reflects the degree of reader demand for the library. However, the number of valid days in the library varies among readers of different grades and majors in universities, so it is necessary to determine individual valid days based on grade and identity. The formula for reader borrowing rate (BR) is Eq. (3.2).

$$BR = \frac{B}{T} \quad (3.2)$$

In equation (3.2), B is the number of times the reader has borrowed. The borrowing behavior of readers also reflects their demand for library book resources. The Utilization rate of Electronic Resources (ER) formula is Eq. (3.3).

$$ER = \frac{\sum_{x \in E} (d_x + l_x)}{T} \quad (3.3)$$

In equation (3.3), E represents the collection of electronic resource databases, x represents a certain database within the collection, d_x represents the download volume in database x , and l_x represents the browse volume. The effective utilization of electronic resources by readers reflects the attractiveness of book resources within the library and reflects the preferences of readers' needs. The formula for the Use of Public Resources (PR) is Eq. (3.4).

$$PR = (pt + zt + kt) / T \quad (3.4)$$

In equation (3.4), pt is the number of times self-service printing is used, zt is the number of times seat reservation is used, and kt is the number of times reading space is used. The number of uses of public resources indicates their attractiveness to readers in order to encourage libraries to increase public resources. The formula for the text features of books borrowed by readers is Eq. (3.5).

$$Z(a_i, y) = \frac{af(a_i, y) \times \log\left(\frac{N}{n_i} + 0.01\right)}{\sqrt{\sum_{a_i \in y} \left[af(a_i, y) \times \log\left(\frac{N}{n_i} + 0.01\right)\right]^2}} \quad (3.5)$$

In equation (3.5), $Z(a_i, y)$ is the weight of feature a_i in all information texts. y is the collection of all textual information. $af(a_i, y)$ is the word frequency of a_i in all texts. N is the total number of information texts. n_i is the number of texts with a_i appearing in the text set, and the denominator is the normalization factor. Vectorization processing is performed on book information, with feature items and their weights forming each dimension of the vector to express readers' needs for the book.

3.2. Design of Library User Profile Module Based on Multi perspective Clustering. The utilization of library book resources not only depends on the book needs of readers, but also on the service quality of librarians in organizing and collecting information, which requires in-depth investigation and analysis of the needs of librarians and readers. The needs of librarians lie in the management of individual portraits of readers and group portraits. The needs of readers include input and modification of personal portrait information, annual report data, personalized recommendation of books, library service recommendation, and interest preference social recommendation.

Using the MD-MBK-means algorithm for clustering a certain dimension of the reader population, and then analyzing the characteristics of several groups to obtain the reader population characteristics. Based on the basic information of readers, a library portrait module for librarians and readers is designed, as displayed in Fig. 3.3.

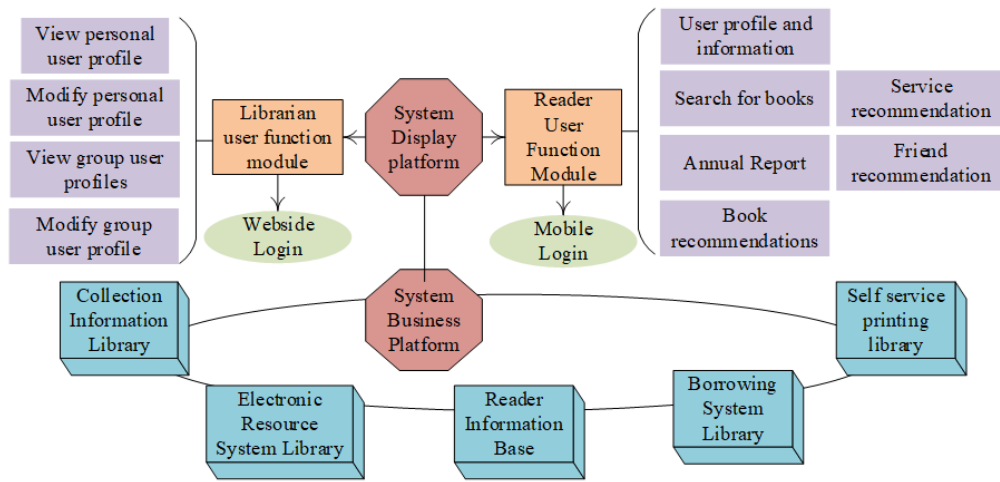


Fig. 3.3: Design of Library User Profile Module

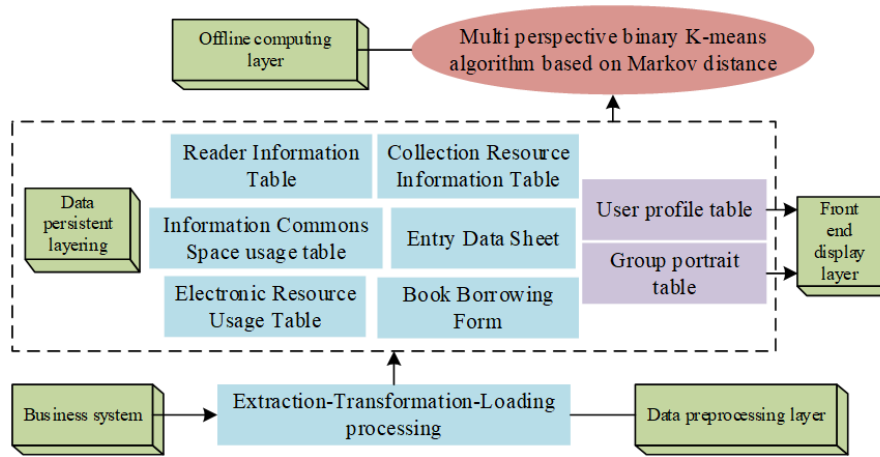


Fig. 3.4: Framework diagram of data computing and backend services

From Fig. 3.3, the business platform in the entire system is most frequently used between librarians and readers, mainly including library collections, electronic resources, reader information databases, borrowing system databases, self-service classical Chinese libraries, and other content. In the front-end platform display, librarians can use web login to view and modify reader user profiles and group user profiles; Readers can use mobile login to modify user profiles, search for book resources, view annual reports, and have functional sections for book services and friend recommendations. The functional modules designed by the system based on the needs of librarians and readers effectively consider their needs and personalized services. Fig. 3.4 shows the framework of the system’s data calculation and backend services.

From Fig. 3.4, the data computing layer adopts MD-MBK-means. The business system used by librarians and readers is preprocessed and inputted into various databases for offline storage of information data. User profiles and group profiles are displayed on the front-end platform system. The clustering analysis used in the system framework is derived from the field of data mining and the improvement of classical K-means. Cluster analysis is the process of aggregating data of the same type, where the differences between the same type of

data are relatively small, while the differences between different types of data are significant. In the process of classifying similar data objects, cluster analysis establishes a similarity matrix based on the similarity of the objects and represents it as a data matrix, as Eq. (3.6).

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} \\ \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} \end{bmatrix} \quad (3.6)$$

In equation (3.6), i represents the number of readers, j represents the attributes of each reader, and x represents the data object. The similarity between different data objects is represented by distance, and the smaller the distance between two objects, the more similar they are. Conversely, the greater the difference. Furthermore, Euclidean distance is used to represent the similarity between two objects, as Eq. (3.7).

$$O(u, v) = \sqrt{\sum_{i=1}^n (u_i - v_i)^2} \quad (3.7)$$

In equation (3.7), O represents the distance between data objects u and v , while u_i and v_i represent the coordinates of u and v , respectively.

3.3. MD-MBK-means algorithm. According to the system framework, clustering algorithms are used to introduce the processes of classic K-means, binary K-means, and MD-MBK-means. Among them, classical K-means is a simple Unsupervised learning method, which can find clusters and cluster centers and is widely used, but it is easy to fall into local optima. Due to the large amount of data on reader behavior characteristics in the study, it is easy to be influenced by the dimension of perspective attributes and optimize the classic K-means. The basic idea of binary K-means is to first receive all datasets into a cluster to form a cluster, recycle one of the clusters, and perform K-means clustering with a cluster number of 2 on it; Then select the two clusters with the smallest total distance from the center of the cluster and place them back in the original cluster set. The cycle will not end until the total number of clusters in the cluster set reaches K. The calculation formula for the cluster center is Eq.(3.8).

$$c = \frac{\sum_{x \in D} x}{m} \quad (3.8)$$

In equation (3.8), c represents the cluster center vector, D represents the dataset, x represents the data objects in the dataset, and m represents the total number of data in the D set. Furthermore, the Sum of Squared Errors (SSE) clustering index is introduced to measure the clustering effect, which calculates the sum of squared Euclidean distances from each data point to the cluster center to obtain the SSE value. The formula is (3.9).

$$SSE_D = \sum_{x \in D} O(x, c)^2 \quad (3.9)$$

In equation (3.9), c represents the center of the dataset D , and $O(x, c)^2$ is the square of the Euclidean distance from x to c . Since Euclidean distance is limited by the dimension of view attribute, which affects the clustering effect and the determination of the optimal cluster number, the calculation formula of MD's Statistical distance is used for multi-attribute analysis. Use the sample matrix X to calculate the formula as Eq.(3.10).

$$M = E\{X\} \quad (3.10)$$

In Formula (3.10), M is the mean value of X , E is the Identity matrix, and the mean value is represented by Identity matrix. The formula for obtaining the mean is (3.11).

$$E\{X\} = X^T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{s \times 1} \quad (3.11)$$

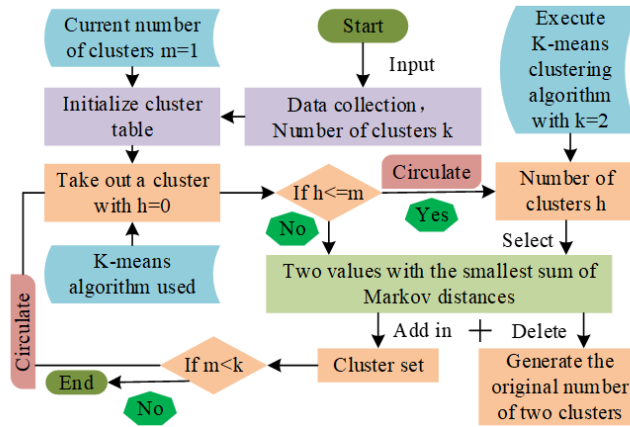


Fig. 3.5: Flow chart of MBK-means algorithm based on MD

In equation (3.11), $(\frac{1}{s})_{s \times 1}$ represents a s -dimensional column vector where all elements are $\frac{1}{s}$. The auto-correlation matrix formula of F is (3.12).

$$C = \frac{X^T \times X}{s} \tag{3.12}$$

In equation (3.12), C represents the autocorrelation matrix. The formula of Covariance matrix L is (3.13).

$$L = E \left\{ (X - M)^T \right\} = \frac{1}{s} X^T X - M \times M^T \tag{3.13}$$

Based on the evaluation results of the sample matrix above, MD will be calculated, as formula (3.14).

$$J^2(X_i - X) = (x_i - M)^T \times \sum_I^{-1} (x_i - M) \tag{3.14}$$

In equation (3.14), X_i and X represent the sample population and sample, respectively, while $J^2(X_i - X)$ represents the Markov distance from the sample to the sample population. MD calculation is based on the overall sample to enhance the clustering effect, and the number of sample data needs to be greater than the sample dimension to ensure the operation of the Covariance matrix. However, the stability of MD clustering algorithm mainly depends on the Covariance matrix and strengthening its stability. Finally, an improved algorithm that replaces Euclidean distance with MD is introduced, and the specific process of combining MBK-means is Fig. 3.5.

From Fig. 3.5, first introducing binary thinking to optimize K-means, and then introducing MD to improve the multi perspective clustering algorithm. Furthermore, the binary K-means is used to classify the number of clusters based on the judgment of the minimum sum of MD. Subsequently, to verify the effectiveness and accuracy of the algorithm, a comparative analysis was conducted on the accuracy of MD-MBK-means, binary K-means, and classical K-means, as Eq. (3.15).

$$A = \sum_{i=1}^k \frac{p_i}{n} \tag{3.15}$$

In equation (3.15), A represents the accuracy of the clustering algorithm, k is the number of clusters, p_i is the number of samples in the i -th individual classification accuracy of the result, and n represents the total number of samples in the sample set.

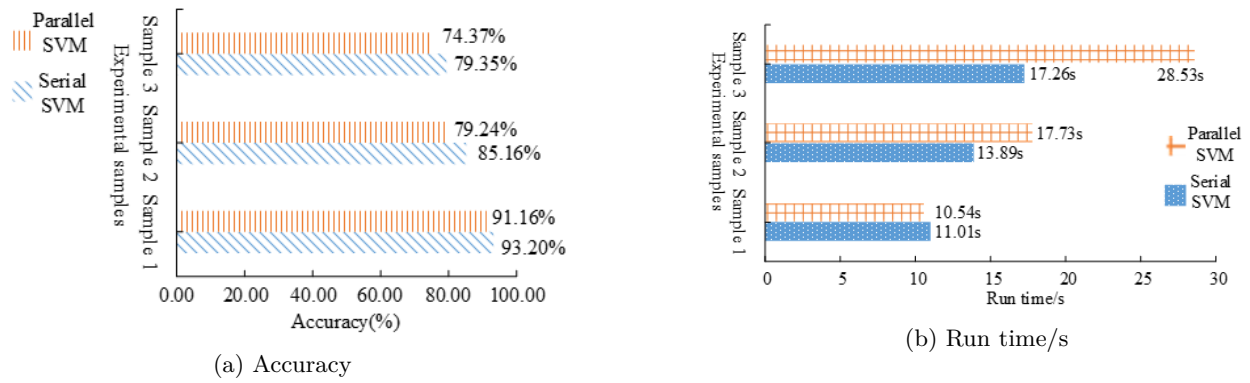


Fig. 4.1: Comparison of accuracy and duration between two algorithms

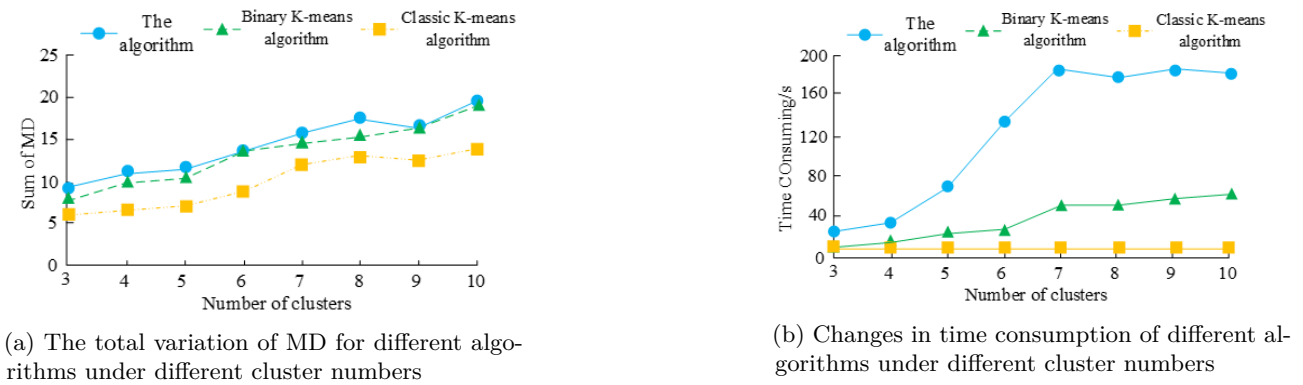


Fig. 4.2: Changes in MD and time consumption of three algorithms under different cluster numbers

4. Comparative Experimental Analysis of MD-MB-K-means Algorithm. One operational method for processing datasets is to parallelize the Spark programming language with Support Vector Machine (SVM) to ensure the efficiency of data mining. The image information of school students is used as experimental subjects and three sets of samples are randomly selected. Then, 70% of each set of samples are set as training samples and 30% as test samples. Subsequently, a Spark based parallel and serial accuracy and duration comparison experiment was conducted to obtain Fig. 4.1.

From Fig. 4.1, the accuracy of parallel SVM in dataset segmentation is 93.20%, 85.16%, and 79.35%, respectively, which are higher than those of serial SVM. The results of parallel SVM in terms of runtime are 11.01 seconds, 13.89 seconds, and 17.26 seconds, respectively, which are generally lower than serial SVM, thus proving the superior performance of parallel SVM in data mining. Then, based on the constructed multidimensional perspective reader feature system, reader data is obtained, and the MD sum and runtime of the clustering results are calculated using MD-MBK-means, binary K-means, and classical K-means, respectively. Fig. 4.2 shows the comparative experimental results.

From Fig. 4.2, the total MD of the MD-MBK-means algorithm is smaller than the binary K-means and classical K-means. Due to the fact that the higher the sum of MD, the better the clustering effect, the MD-MBK-means has the best clustering effect and can avoid falling into local optima. In the time variation experiment, the same algorithm was performed 10 times under each cluster number to obtain the MD-MBK-means with the

Table 4.1: Number, proportion and range of indicators for each group

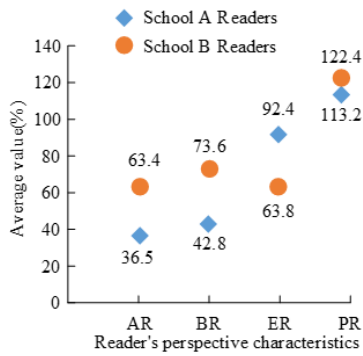
Group	Number	Proportion	AR range	BR range	ER range	PR range
Group 1	9847	51.14%	[0,0.236]	[0,0.461]	[0,0.738]	[1.253,2.064]
Group 2	3169	16.46%	[0.454,1.125]	[0.496,1.022]	[0.675,1.285]	[0,0.497]
Group 3	160	0.83%	[1.582,3.580]	[0.623,0.723]	[1.758,2.527]	[0.365,0.582]
Group 4	4536	23.56%	[0.237,0.545]	[0.782,0.935]	[1.264,1.862]	[1.165,1.348]
Group 5	1544	8.06%	[0.850,1.768]	[0.378,0.823]	[2.235,5.492]	[0.462,1.254]

shortest time, highest efficiency, and stable time consumption. Therefore, MD-MBK-means outperforms the other two algorithms in clustering performance, time consumption, and stability. Then, based on the algorithm indicators of the reader feature system, the reader perspectives in the system are classified. A K-5 clustering analysis is conducted on 19,256 readers from a certain university to obtain 5 user groups, and the number of readers and the proportion of each indicator are calculated. As listed in Table 4.1.

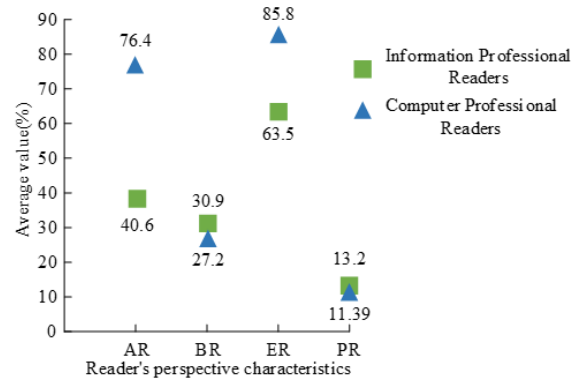
Table 4.1 shows the proportion of people in each reader group and the range of perspectives, and grading different ranges to distinguish between high and low is necessary to obtain user group characteristics for the system settings. Customized reader activity, reader borrowing rate The highest and lowest levels of electronic resource utilization and public resource utilization from the four reader perspectives are ≥ 1.5 and $[0,0.5]$, ≥ 0.8 and $[0,0.3]$, ≥ 1.6 and $[0,0.7]$, and ≥ 1.0 and $[0,0.4]$. The results can customize personalized services for readers and users to achieve precise services and personalized recommendations. Then, multi-dimensional portrait comparisons of readers and users are conducted, and different dimensions are used to classify readers and analyze the characteristics of books borrowed by readers to achieve targeted and personalized services. The experiment uses four dimensions: school, major, gender, and grade to compare and summarize the characteristics of readers' perspectives, as Fig. 4.3.

In Fig 4.3, the reader activity, borrowing rate, and electronic resource utilization rate of the 19256 reader groups in School A are low, while the high utilization rate of public resources indicates that the service plan for the school is to recommend public and electronic resources. The reader activity and electronic resource utilization rate of 21485 readers in School B are low, while the borrowing rate and public resource utilization rate are high, indicating that their service recommendations are based on the collection resources and public resources. In the comparison chart of the professional dimension, it is found that information and computer readers have lower results in terms of reader activity, borrowing rate, and electronic resource utilization, while the utilization rate of public resources is also higher. However, the recommended services vary depending on the profession and the characteristics of the books used. The former is mainly recommended for public resources and electronic resources, while the latter is mainly recommended for library activities and electronic resources. Then, by comparing gender, it can be concluded that service recommendations for male readers are mainly electronic and public resources, while service recommendations for female readers are library activities and collections of books. Finally, comparing the grade dimensions, it was found that the service recommendations of 2981 readers in the 2016 grade mainly focused on electronic resources and public resources; In 2015, 4004 readers recommended electronic and public resources. Afterwards, the constructed multi-dimensional reader feature system will be used for MD-MBK-means and clustering experimental analysis. To verify the feasibility of this algorithm, it was compared and analyzed with binary K-means and classical K-means in terms of algorithm accuracy, MD sum, and algorithm time. Fig. 4.4 shows the comparison of three algorithms under the algorithm accuracy formula.

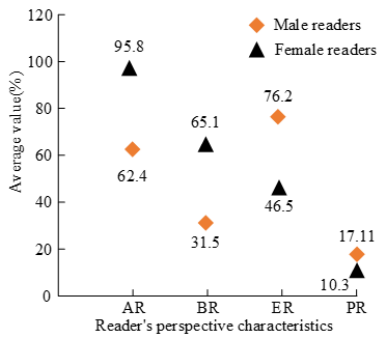
The datasets used for comparison in Fig. 4.4 are Iris Data Set (Iris), Liver injuries Data Set (Liver), and Ionosphere Data Set (Ionosphere). The result shows that the accuracy of MD-MBK-means is the highest in all three sample sets, with values of 92.56%, 78.62%, and 72.57%, respectively, and they are all higher than classical K-means and binary K-means. All three datasets belong to machine learning, among which the Iris dataset, as a relatively old dataset, is often used to introduce linear discriminant analysis and contains 150 samples. The sample features are the sepal length and width of iris flowers and the petal length and width. The Liver dataset mainly includes liver injury and its multivariate, standard test datasets, while the Ionosphere dataset includes



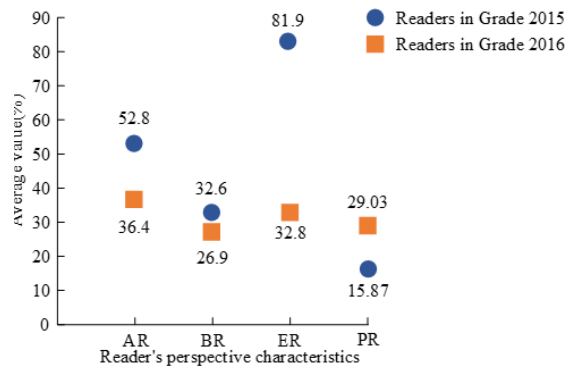
(a) Comparison Table of School Dimensional Perspective Characteristics



(b) Comparison of Professional Perspective Characteristics



(c) Comparison of Gender Dimension Perspective Characteristics



(d) Comparison of Perspective Characteristics of Grade Dimensions

Fig. 4.3: Comparison of Different Dimensions on the Characteristics of Reader's Perspectives

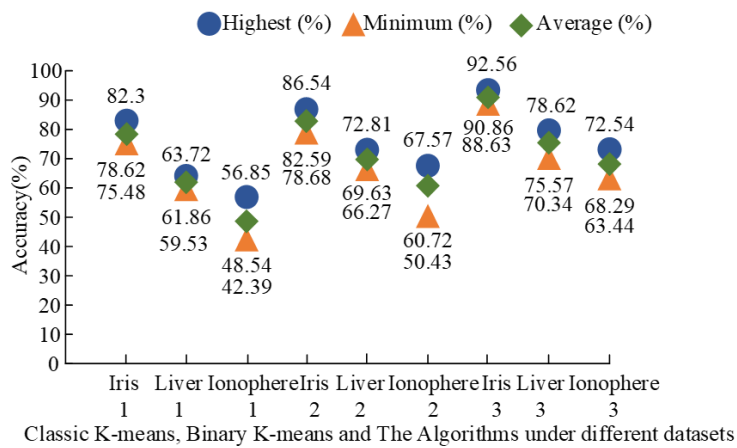


Fig. 4.4: Comparison of Accuracy of Three Algorithms

Table 4.2: Comparison of Clustering Effects of three algorithms on different sample sets

Algorithm	Sample set	Highest (Sum of MD)	Minimum (Sum of MD)	Average (Sum of MD)
Classic K-means algorithm	Iris	5.526	3.649	4.400
	Liver	7.687	6.398	7.159
	Ionosphere	38.486	30.726	33.682
Binary K-means algorithm	Iris	4.173	4.136	4.158
	Liver	6.872	6.136	6.652
	Ionosphere	35.549	28.527	30.296
The algorithm	Iris	3.854	3.543	3.685
	Liver	6.357	5.268	5.872
	Ionosphere	24.861	22.385	23.026

Table 4.3: Comparison of Time Consumption of three algorithms for each sample set

Algorithm	Sample set	Highest (Second)	Minimum (Second)	Average (Second)
Classic K-means algorithm	Iris	0.036	0.024	0.031
	Liver	0.034	0.094	0.065
	Ionosphere	0.051	0.113	0.075
Binary K-means algorithm	Iris	0.033	0.027	0.030
	Liver	0.035	0.080	0.059
	Ionosphere	0.038	0.099	0.074
The algorithm	Iris	0.030	0.021	0.027
	Liver	0.029	0.056	0.048
	Ionosphere	0.031	0.085	0.062

ionosphere and its classification, multivariate, and standard test datasets. Compare the MD sum of the three algorithms again to solve the defect of local optimum in classical K-means and obtain Table 4.2.

From Table 4.2, the smaller the sum of MD values, the better the clustering effect. The sum of MD of MD-MBK-means in the three sample sets is smaller than that of classical K-means and binary K-means, with values of 3.543, 5.268, and 22.385, which proves the advantage of its algorithm clustering performance. Finally, the running time of the three algorithms was compared to consider computational efficiency, as Table 4.3.

In Table 4.3, the shorter the algorithm running time, the higher the computational efficiency. The results show that the running time of MD-MBK-means in the three sample sets is 0.021s, 0.056s, and 0.085s, and is significantly better than classical K-means and binary K-means.

5. Conclusion. Regarding the design of LRPT and its system, this study proposes to partition reader information and analyze its perspective characteristics based on the multidimensional perspective feature system of readers. Firstly, based on the design framework of the portrait system, MBK-means is introduced to optimize the classic K-means. Secondly, the calculation results of MD sum are compared between MD-MBK-means, binary K-means, and classic K-means. The experiment shows that the MD sum of MD-MBK-means is the smallest and the clustering effect is the best. This study also compared the multi-dimensional aspects of reader perspective characteristics and recommended services to various dimensions of reader groups based on the personalized service function of the system. In the final experiment, three sample sets were used to compare the accuracy, MD sum, and running time of the three algorithms. Among the three sample sets, the MD multi MBK means algorithm had the highest accuracy, with values of 92.56%, 78.62%, and 72.57%; The sum of MD results is the smallest, with values of 3.543, 5.268, and 22.385, respectively; The running time is the shortest, with a duration of 0.021s, 0.056s, and 0.085s for the three sample sets. The multi perspective clustering algorithm provides rich and accurate algorithms for the customization needs of library reader profiles. The reader profile system also deepens the correlation between readers and book resources, and enhances readers' reading interest. This proves the superiority of the MD multi perspective binary K-means algorithm. However,

its algorithm still lacks experimental data on the dimensional features of reader portrait perspectives, as well as the representativeness of extensive sample set experimental data. Therefore, further research and improvement are needed in future research.

REFERENCES

- [1] Byrkovych, T., Humenchuk, A., Kobyzcha, N., Akimova, L. & Grinberg, L. Economic Effectiveness of State Policy in Reforming Higher Library and Information Education in Ukraine. *Economic Affairs*. **68**, 599-616 (2023)
- [2] Wangdi, S., LeGrand, C., Norbu, P. & Rinzin, S. What's past is prologue: history, current status and future prospects of library development in Bhutan. *Global Knowledge, Memory And Communication*. **70**, 339-354 (2021)
- [3] Van Melik, R. Merry M S. Retooling the public library as social infrastructure: a Dutch illustration. *Social & Cultural Geography*. **24**, 758-777 (2023)
- [4] Melnikova, N. & Romanovskaya, O. The Russian concept of social informatics in light of information technology innovation: a systematic review. *Acta Informatica Pragensia*. **10**, 301-332 (2021)
- [5] Zhou, L. Research on user portrait data management strategy of mobile library based on blockchain technology. *Library Work And Research*. (7) pp. 49-57 (2021)
- [6] Navarrete, T. & Villaespesa, E. Image-based information: paintings in Wikipedia. *Journal Of Documentation*. **77**, 359-380 (2021)
- [7] Hamilton, J., Stapleton, B. & Plaisance, H. More than just a walk through: Connect library users to resources with new 360 tools. *College & Undergraduate Libraries*. **27**, 176-196 (2021)
- [8] Zeng, Z., Sun, S., Li, T., Yin, D. & Shen, Y. Mobile visual search model for Dunhuang murals in the smart library. *Library Hi Tech*. **40**, 1796-1818 (2022)
- [9] Chen, Z., Zhou, M. & Feng, L. Analysis of the smart library construction in colleges based big data and artificial intelligence. *Journal Of Physics: Conference Series. IOP Publishing*. **1955**, 12-17 (2021)
- [10] Qu, J. Research on mobile learning in a teaching information service system based on a big data driven environment. *Education And Information Technologies*. **26**, 6183-6201 (2021)
- [11] Ahmad, R. & Rafiq, M. Assessing the preparedness of university libraries for digital preservation. *The Journal Of Academic Librarianship*. **48**, 102-617 (2022)
- [12] Shiri, A., Howard, D. & Farnel, S. Indigenous digital storytelling: digital interfaces supporting cultural heritage preservation and access. *International Information & Library Review*. **54**, 93-114 (2022)
- [13] Zhang, L., Zhao, L., Ren, H., Xiao, Y., Ma, Y. & Zhang, Q. Knowledge graph and behavior portrait of intelligent attack against path planning. *International Journal Of Intelligent Systems*. **37**, 7110-7123 (2022)
- [14] Li, W. Design of smart campus management system based on internet of things technology. *Journal Of Intelligent & Fuzzy Systems*. **40**, 3159-3168 (2021)
- [15] Candela, G., Sáez, M., Escobar Esteban, M. & Marco-Such, M. Reusing digital collections from GLAM institutions. *Journal Of Information Science*. **48**, 251-267 (2022)
- [16] Wang, F., Zhang, L., Chen, X. & Others A personalized self-learning system based on knowledge graph and differential evolution algorithm. *Concurrency And Computation: Practice And Experience*. **34**, 61-90 (2022)
- [17] Tong, L., Zhang, C. & Huang, R. Research on intelligent logic design and application of campus MMTC scene based on 5G slicing technology. *China Communications*. **18**, 307-315 (2021)
- [18] Gao, T. Research on short video recommendation strategy based on big data analysis. *Journal Of Physics: Conference Series. IOP Publishing*. **1941**, 12-071 (2021)
- [19] Long, X., Chen, Y. & Zhou, J. Development of AR Experiment on Electric-Thermal Effect by Open Framework with Simulation-Based Asset and User-Defined Input. *Artificial Intelligence And Applications*. **1**, 52-57 (2023)

Edited by: Mudasir Mohd

Special issue on: Scalable Computing in Online and Blended Learning Environments: Challenges and Solutions

Received: Nov 13, 2023

Accepted: Apr 1, 2024