Factors and Models Influencing Value Co-Creation in the Supply Chain of Collection Resources for Library Distribution Providers Under Data Ecology

Xiaoyun Lin*
Library
Guangzhou College of Technology and Business
Guangzhou, 510850, China

***Abstract*—Under the data ecology, the advancement of relevant technology and the utilization of relevant resources have provided more efficient technical services for various industries. However, with the proliferation of data resources, problems such as information pollution and data redundancy have arisen in the process of supply chain services for collection resources. For solving such problems and enhancing the collection resource supply efficiency of librarians, the study uses data mining technology combined with improved K-Means clustering algorithm to design a value co-creation model of library collection resource supply chain for librarians under data ecology. The outcomes indicate that the shortest running time of traditional K-Means algorithm is 40ms and the longest running time is 115ms in Wine dataset, and the running time of improved K-Means algorithm is stable at 59ms; the shortest running time of traditional K-Means algorithm is 26 ms and the longest running time is 58ms in Iris dataset, and the running time of improved K-Means algorithm is stable at 53 ms. The clustering accuracy of the improved K-Means algorithm in the Wine data set is 98.2%, which is 0.3% exceeding the traditional K-Means algorithm, which is 97.9%; the clustering accuracy in the Iris data set is 100%, which is 2.4% exceeding the traditional K-Means algorithm, which is 97.6%. In summary, it can be seen that the studied data ecology has a good application of the factors and models influencing the value co-creation of the supply chain of library resources for library dispensers.**

Keywords—Resource supply chain; data mining; value co-creation; K-Means clustering algorithm; pavilion dispenser

I. Introduction

With the big data's reach, the use of data resources changed the current human lifestyle and cognitive level, and data resources have become an important asset [1]. The advancement of relevant technology has led to the increasing participation of users in the network, and the phenomenon that everyone has data but everyone lacks data has emerged in various industries [2]. The increase in data transparency and freedom and the increasing circulation of data have led to information pollution and data redundancy in the current data ecosystem [3]. Therefore, the traditional library resource management model cannot satisfy the needs of efficient resource integration for the massive library collection resources nowadays. Data mining (DM) is a data-centered mode of thinking, which can rely on algorithms to effectively refine and integrate generous data to maximize the benefits of data resources [4]. The integration of resources based on DM technology is the service goal of the supply chain of collection resources of the librarians, and the deep combination of DM and supply chain resource management to explore a high-efficiency and low-cost resource supply chain management model is the method for realizing the relevant value of collection resources. The K-Means (KM) algorithm is extensively utilized classical classification clustering algorithms (CA). It can effectively solve the low clustering accuracy and simple for falling into local solutions [5]. The problem and goal of this article are to improve the efficiency of collection resource supply for library distributors. The significance and contribution of this study are to propose the connotation of value co-creation (VCC) in the collection resource supply chain of library distributors, provide readers with a better knowledge service environment, and provide theoretical basis and direction guidance for future research on the collection resource supply chain. Therefore, in this context, the factors and models influencing the VCC in the supply chain of library resources for library distributors under the data ecology are studied to improve the use rate and realize the VCC in the supply chain of library resources. The second part is a review of the research of resource supply chain and DM technology at home and abroad; the third part is a study on the VCC model of library resource supply chain based on the data ecology. The first section analyzes the factors affecting the VCC of library resource supply chain based on the Decision-making Trial and Evaluation Laboratory (DEMATEL)_ method; the second section is the constructing of the VCC model of library resource supply chain for library distributors. The fourth part is the validation of the VCC model of the supply chain of library resources for library service providers based on the data ecology.

II. Related Works

The resource distribution link of the resource supply chain introduces new-age technologies such as logistics technology and Internet of Things. Therefore, the resource supply chain based on data and emerging technologies has gained the attention of many domestic and foreign experts and scholars, and researchers have carried related research on it. Nandi M L and other scholars proposed integrating relevant technology into its corresponding supply chain system to improve the supply chain performance and construct a resource-based supply chain system framework. It is based on 126 secondary information cases and a qualitative content analysis of these 126 supply chain system cases using a retrospective approach. The outcomes illustrated that the method is useful in improving information sharing in supply chain and has reference value in supply chain integration [6]. The research team of Agyabeng-Mensah Y, to test the influence of green human resource management on internal green supply chain, proposed an equation modeling software with partial least squares structure, which utilized a relevant method to collect data in supply chain and used partial least squares structure for analyzing the data. The outcomes showed that the method tested the internal green supply chain and was feasible [7]. Mehrotra S et al. proposed to establish a stochastic optimal allocation model with sharing function for enhancing the system of the ventilator supply chain, which was mainly used for the allocation of ventilator resources during the new coronary pneumonia pandemic. The outcomes indicated it could enhance the allocation efficiency of the ventilator supply chain [8].

DM techniques possess an essential influence in resource supply chain research. Zhao Y's research team proposed a building energy fault diagnosis model based on DM techniques for enhancing the operation of energy systems in the building industry. Then the model utilized unsupervised DM techniques to construct a framework for building energy operation pattern identification and energy load diagnosis, and the model was tested in a dataset Validation. The outcomes indicated that the model possessed excellent accuracy of building energy load prediction [9]. Liu J and other scholars proposed a privacy-preserving DM algorithm in view of Bayesian analysis and logical analysis for addressing the data privacy protection. The core of the method was for letting DM technology become data and strengthen the management of DM technology. The outcomes indicated that the method improved the security of privacy protection by 20% and is feasible [10]. Rong Z and other researchers proposed a multi-layer fuzzy evaluation model in view of AI DM, which integrated fuzzy information such as learning, recognition, correlation and adaptive to evaluate the ideological education of undergraduates, in response to the problems of unclear evaluation system and single evaluation method of undergraduates. The outcomes illustrated that the model achieved the desired effect with scientific validity [11].

In summary, DM techniques are often applied in supply chain research in various fields, but research on using big DM techniques to build a VCC model for the supply chain of library resources is quite rare. In this context, the factors and models influencing the VCC of the supply chain of collection resources for library dispensers under the data ecology are studied to achieve better application results.

III. Research on the VCC of the Supply Chain of Library Resources in View of the Data Ecology of Library Dispensers

In this chapter, the structure of the supply chain of library resources of library dispensers under the data ecology is studied, and the factors influencing the VCC of the supply chain of library resources are analyzed using the DEMATEL method, and the construction of the VCC model of the supply chain of library resources of library dispensers is carried out using the improved KM CA.

A. *Analysis of the Factors Influencing the VCC in the Supply Chain of Collection Resources based on the DEMATEL Method*

Data ecology is evolved from the digitization of knowledge resources, which is generated as an outcome of the combination of resource digitization and information technology (IT). Data ecology is based on big data carriers, forming a community of interest for data sharing and collaboration to achieve the purpose of VCC [12]. From the perspective of supply chain, libraries can be regarded as logistics distribution centers, and the resource supply chain links upstream collection suppliers and downstream library customers. The theory of VCC in the supply chain of collection resources can be used in the library distribution industry, where libraries and distribution providers work together to provide quality collection resources to library customers and create social value to improve social and economic benefits [13]. The supply chain structure of library resources for library distributors under data ecology is shown in Fig. 1.

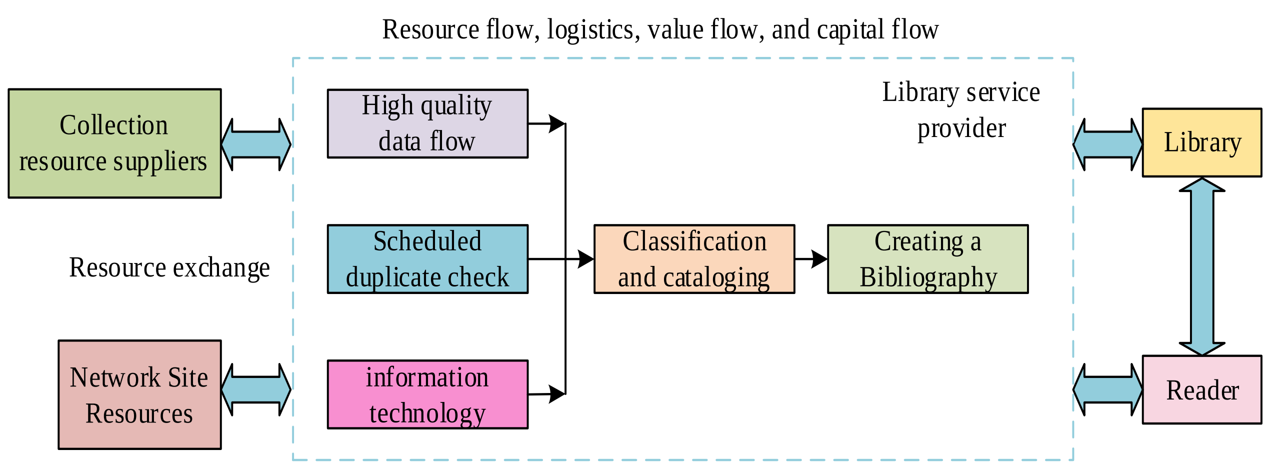


Fig. 1. The supply chain structure of collection resources for library distributors in the data ecology.

As shown in Fig. 1, the major part of the supply chain structure of collection resources mainly contains libraries, library distributors, library customers and collection suppliers. The collection resources mainly include book resources, e-book resources and library intelligent auxiliary devices provided by the collection suppliers. To realize the VCC in the supply chain of collection resources, it is essential for bringing into play the supply efficiency of collection resources and ensures the effective circulation of resource flow, capital flow, logistics and value flow in the supply chain of collection resources [14]. According to the root theory, nine factors can be derived to influence the supply chain VCC: the construction and integration of supply chain resources, the advancement and applying of supply chain service technology, the applying of data for supply chain VCC, the quality of staff in the supply chain, the concept of supply chain VCC, the construction of supply chain VCC platform, the interaction of the main body of supply chain VCC, the demand of readers for supply chain VCC, and the environmental constraints of supply chain VCC, etc. [15]. DEMATEL is a way of systematic analysis using matrix tools and graph theory, which can calculate the influence degree of each factor based on the logical relation in the elements [16]. For identifying the key factors in the supply chain VCC from the nine factors, the degree of correlation between the factors is calculated using the DEMATEL method. The mathematical expression of the normalization matrix is shown in Eq. (1).

$$H = \frac{X}{\max\limits_{1 \le i \le n} \sum_{j=1}^{n} x_{ij}} \tag{1}$$

In Eq. (1), X denotes the influence relationship matrix while H means the normalized one. x_{ij} denotes the degree of influence of factor i on factor j. The relevant formula is demonstrated in Eq. (2).

$$S = \lim_{k \to \infty}\left(H + H^2 + \ldots + H^k\right) = \sum_{k=1}^{\infty} H^k = H\left(K - H\right)^{-1} \tag{2}$$

In Eq. (2), K denotes the unit matrix and S indicates the combined influence matrix. The method for calculating the influence degree of each factor is shown in Eq. (3).

$$Q = \sum_{j=1}^{n} s_{ij} \quad \left(i = 1, 2, \ldots, N\right) \tag{3}$$

In Eq. (3), Q denotes the influence degree vector of factors and N expresses the total of factors. The influenced degree vector is shown in (4).

$$W = \sum_{i=1}^{n} s_{ij} \quad \left(j = 1, 2, \ldots, N\right) \tag{4}$$

In Eq. (4), W indicates the vector of factors' influence degree, and the centrality and the reason degree can be calculated according to the influence degree of the factors. Reason degree illustrates the influence of the factor on other factors, and a positive reason degree illustrates that the element possesses an impact on other elements, while a negative reason degree illustrates that the element is affected by other elements [17]. The relevant formula is showcased in Eq. (5).

$$U_i = Q_i + W_i \tag{5}$$

In Eq. (5), U indicates the value of centrality. The relevant formula is showcased in Eq. (6)

$$V_i = Q_i - W_i \tag{6}$$

In Eq. (6), V indicates the value of the cause degree. The combined importance of the system elements can be counted from the values of centrality and causality, as shown in Eq. (7).

$$Z_i = \alpha_i U_i + \beta_i V_i \tag{7}$$

In Eq. (7), Z_i represents the combined importance of system factors; α denotes the pendulum weighting coefficient of centrality, and β refers to the pendulum weighting coefficient of cause. Based on the maximum and minimum values of the centrality and the reason degree, the optimal and the worst solutions can be derived, and then the swing weight coefficients are used to assign weights to the indicators of the two solutions. Then the final weight values of the centrality and the reason degree can be obtained by normalizing the two assigned weights. Based on the key factors identified by the DEMATEL method, a model of co-creation influence factors of the supply chain value of the collection resources based on the DEMATEL method is designed as shown in Fig. 2.

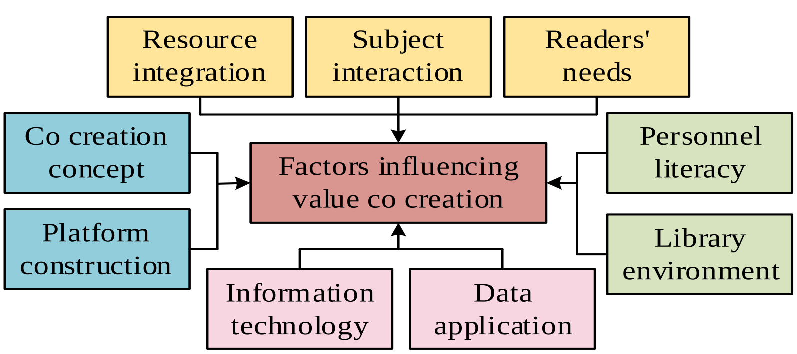


Fig. 2. A model of factors influencing the VCC of collection resources supply chain based on DEMATEL method.

As shown in Fig. 2, in the VCC influence factor model, platform construction and co-creation concept are the key factors of supply chain VCC of library resources; readers' demand, subject interaction and resource integration are the drivers of supply chain VCC; platform environment and staff literacy are the direct factors of supply chain VCC; and data applying and IT are the indirect factors of supply chain VCC.

B. The Construction of the VCC Model of the Supply Chain of Library Resources for Library Distribution Providers

The realizing of VCC in the supply chain of collection resources under the data ecology cannot be achieved without the influence of nine factors, while the complexity of the operation and management of the supply chain of collection resources brings challenges to the realization of VCC. For reaching the purpose of providing readers with quality resources in the resource supply chain, DM technology is utilized for exploring the VCC in the collection resource supply chain and build a model of VCC in the collection resource supply chain with DM. DM is a data-centered mode of thinking, and DM relies on algorithms and technologies to effectively refine and integrate massive amounts of data to maximize the benefits of data resources [18]. For achieving the VCC process of the supply chain of collection resources under the data ecology, the VCC model of the supply chain of collection resources for library distribution providers is constructed as shown in Fig. 3.

As shown in Fig. 3, the data base in the VCC model of the supply chain of library resources contains equipment operation data, organization operation data, reader behavior data and resource integration data. Effective DM of the collection resources under the big data ecology is beneficial for the library distributor to understand the information of readers' needs and library business weaknesses, and it can also enhance the communication efficiency among the supply chain subjects. The core idea of DM technology is mining potentially useful information from generous disordered and incomplete data using relevant algorithms. The relevant basic process is showcased in Fig. 4.

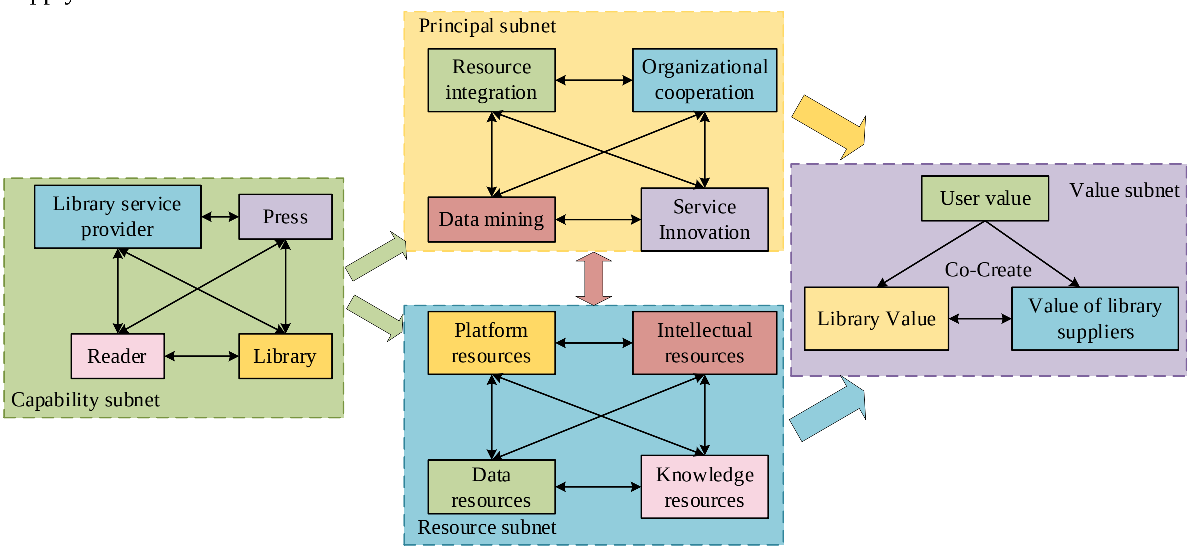


Fig. 3. The VCC model of collection resources supply chain for library distributors.

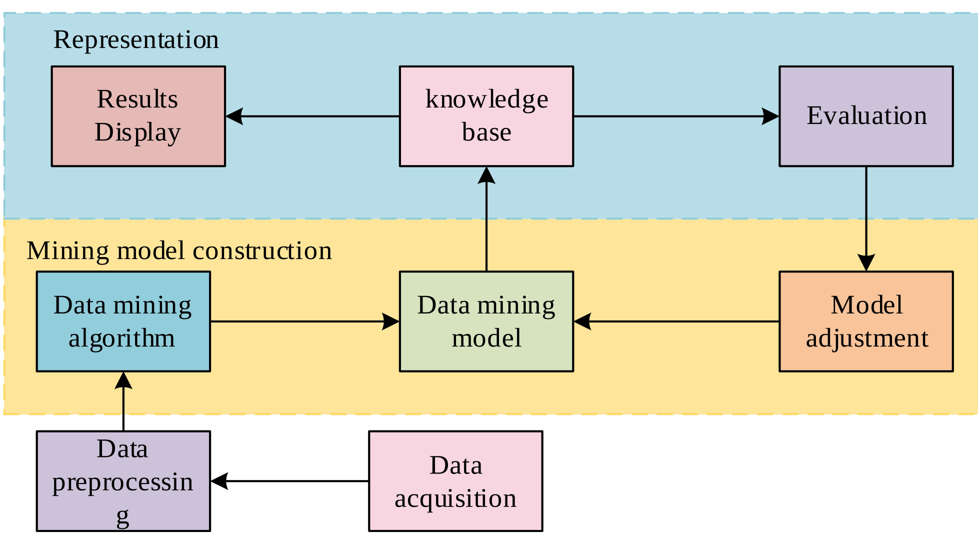


Fig. 4. The basic process of DM.

Fig. 4 indicates that the DM includes steps of data collection, data pre-processing, construction of DM models and output of results. The algorithms for building DM models are generally divided into four categories, such as CA, association rule algorithms, classification and regression algorithms, and CA [19]. The KM algorithm is extensively utilized CA as its simple principle and easy implementation [20]. The CA is an indirect clustering division method, which can measure the similarity between samples using Euclidean distance (ED) and transform it by means of data distance metric for better calculation of data. The CA utilizes distance as a criterion for evaluating the metric of sample similarity. The closer the distance in two samples, the higher the similarity between the samples. The most used distance formulas in CA are ED and Manhattan distance, and the mathematical expression of ED is indicated in Eq. (8).

$$D(i,j)=\sqrt{\left(x_{i1}-x_{j1}\right)^2+\left(x_{i2}-x_{j2}\right)^2+...,+\left(x_{in}-x_{jn}\right)^2} \tag{8}$$

In Eq. (8), D denotes the ED, i and j denote the data objects, and n denotes the dimensionality. The mathematical expression of the Manhattan distance is shown in Eq. (9).

$$D'(i,j)=\left|x_{i1}-x_{j1}\right|+\left|x_{i2}-x_{j2}\right|+...,+\left|x_{in}-x_{jn}\right| \tag{9}$$

In Eq. (9), D' denotes the Manhattan distance. If no end criterion is set in the KMA, the distance calculation of Eq. (8) and Eq. (9) will be repeated, and reaching the end mark means that the algorithm has reached convergence. The flow chart of the KMA is illustrated in Fig. 5.

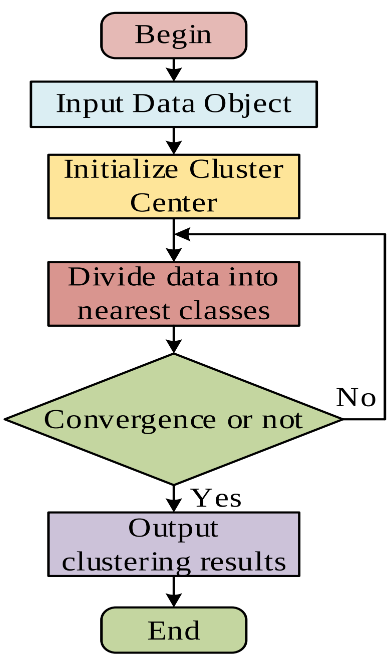


Fig. 5. Flowchart of KM algorithm.

Fig. 5 indicates that the algorithm first prepares the data, determines the number of classifications and sets the clustering centers, then counts the range in each data point and the clustering center, and groups the closest points into the initial centroids. After that, the clustering centroids are recalculated and the convergence is judged. If the convergence is not reached, the above steps are repeated until convergence is achieved. The most common method used in CA to determine whether convergence is achieved is the error-sum-squared criterion method, and the mathematical expression of the error-sum-squared criterion method is shown in Eq. (10).

$$J_D=\sum_{i=1}^{D}\sum_{k=1}^{m_i}\left\|x_k-m_i\right\|^2 \tag{10}$$

In Eq. (10), D denotes the sample set; J denotes the objective function; m_i serves as the mean value of the sample, and the formula for calculating the sample mean is shown in Eq. (11).

$$m_i=\frac{1}{n_i}\sum_{i=1}^{n_i}x_i \quad i=1,2,...,D \tag{11}$$

According to Eq. (10) and Eq. (11), the value of the objective function depends on the sample centroid, and the larger the sample centroid, the larger the objective function and consequently the larger the error; conversely, the smaller the sample centroid, the smaller the objective function, and consequently the smaller the error. For reducing the probability of error in the calculation, the traditional KM algorithm is enhanced by using the weighted squared average distance method to reduce the search range in the calculation. The expression of the weighted squared average distance method is shown in Eq. (12).

$$J_i=\sum_{i=1}^{D}P_iA_i^* \tag{12}$$

In Eq. (12), A_i^* denotes the mean squared distance between samples and P_i denotes the prior probability. The formula for calculating the mean squared distance is shown in Eq. (13).

$$A_i^*=\frac{2\sum_{x\in X_i}\sum_{x\in X_i}\left\|x-x'\right\|^2}{n_i\left(n_i-1\right)} \tag{13}$$

In Eq. (13), n_i denotes the number of data samples, and $\sum_{x\in X_i}\sum_{x\in X_i}\left\|x-x'\right\|^2$ denotes the sum of distances between samples. The mathematical expression of the interclass distance and criterion is shown in Eq. (14).

$$J_{\delta 1}=\sum_{i=1}^{D}\left(m_i-m\right)^T\left(m_i-m\right) \tag{14}$$

In Eq. (14), m denotes the mean value of the distance between samples, and m_i denotes the mean value of the distance between samples of i. The weighted interclass distances and criteria are shown in Eq. (15).

$$J_{\delta 2} = \sum_{i=1}^{D} P_i \left(m_i - m\right)^T \left(m_i - m\right) \tag{15}$$

As shown by Eq. (15), the weighted interclass distance and criterion is the association of adding a priori probability to the interclass distance and criterion formula. The larger the value of J_δ corresponds to the greater the degree of variation, then the more obvious the clustering analysis results.

IV. Validation of the VCC of the Supply Chain of Library Resources in View of the Data Ecology of Library Distributors

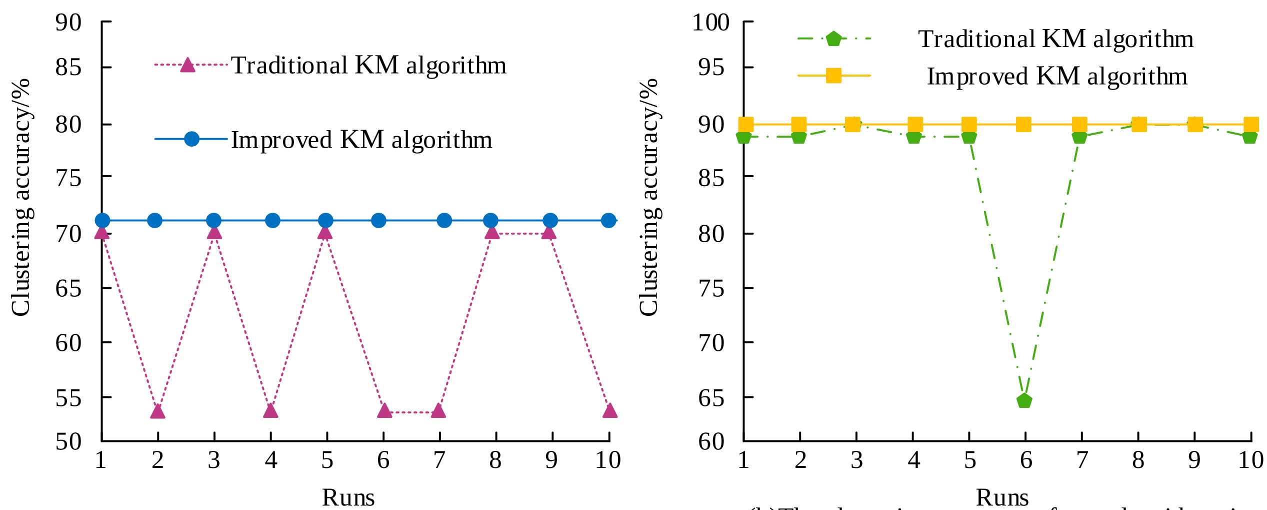
For verifying the performance of the VCC model of the supply chain of library resources based on the data ecology, the performance of the traditional KM algorithm was compared with that of the improved one. The total samples in the Wine dataset were 440, and there were four clusters of the same size and no isolated points in the dataset; the total number of samples in the Iris dataset was 300, and there were three clusters of the same size and five isolated points in the dataset. The specific the relevant outcomes are showcased in Table I.

The traditional and the improved KM algorithms were run ten times in each of the two training sets, and the comparing of the CA of the two algorithms run ten times in the data set is shown in Fig. 6. Fig. 6(a) illustrated that the traditional KM algorithm showed an up-and-down state in the Wine dataset, with the highest CA of 70.3% and the lowest of 53.5%, while the CA of the improved KM algorithm was stable at 71%. In Fig. 6(b), the traditional KM algorithm achieved the lowest CA of 65% at the sixth run in the Iris dataset, and the CA fluctuated around 89% for the other runs, while the CA of the improved KM algorithm remained at 90%. Collectively, it illustrated that the traditional KM algorithm was not stable enough and the improved KM algorithm had more stability.

For comparing the performance of the traditional and the improved KM algorithms more comprehensively, the running times of the two algorithms were studied. The running time comparison of the two algorithms run ten times in the dataset is indicated in Fig. 7. Fig. 7(a) showcased that the running time of the traditional KM algorithm in the Wine dataset still presented an unstable state, in which the shortest running time was 40 ms and the longest running time was 115 ms, while the running time of the improved KM algorithm was stable at 59 ms. Fig. 7(b) indicated that the traditional KM algorithm in the Iris dataset had the shortest running time of 40 ms and the longest running time of 115 ms, while the running time of the improved KM algorithm was stable at 59 ms. The shortest running time of the traditional KM algorithm was 26ms and the longest running time was 58ms, while the running time of the improved KM algorithm stayed at 53ms. Collectively, it illustrated that the improved KM algorithm had stronger robustness compared to the traditional KM algorithm.

TABLE I. EXPERIMENTAL ENVIRONMENT

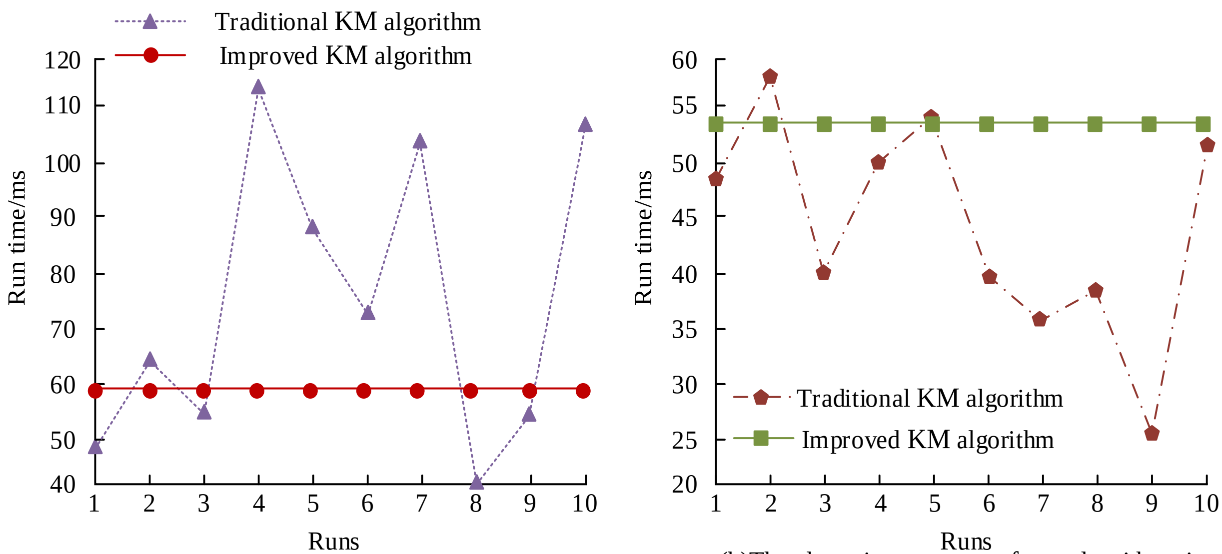
Experimental Environment	Configuration
Operating system	Windows 10
Memory	16G
GPU	NVIDIA TITAN BLACK GPU
CPU	Intel(R) Core (TM) i5-4460
Programming Language	Python
Algorithm testing tools	PyCharm



(a)The clustering accuracy of two algorithms in dataset Wine

(b)The clustering accuracy of two algorithms in dataset Iris

Fig. 6. Comparison of CA between two algorithms running ten times.



(a)The clustering accuracy of two algorithms in dataset Wine

(b)The clustering accuracy of two algorithms in dataset Iris

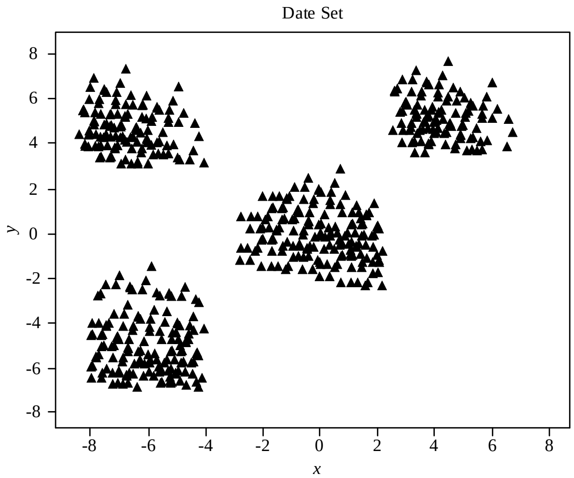
Fig. 7. Comparison of the runtime of two algorithms in the dataset.

To observe more intuitively the performance comparing between the traditional and the improved KM algorithms, the performance metrics of the two algorithms were compared and analyzed with the number of iterations. The comparison results are showcased in Table II. From Table II, the average CA of the improved KM algorithm was 71% and 90% in the Wine dataset and Iris dataset, respectively, which was 9.1% and 3.4% higher than the average CA of the traditional KM algorithm. And it showcased that the improved KM algorithm had higher CA, and the improved KM algorithm only runs the best clustering effect could be achieved by running only once. In terms of the number of iterations, that of the improved KM algorithm was smaller than the average iteration times of the traditional KM algorithm. Although the average running time of the improved KM algorithm was 53 ms in the Iris dataset, which was 8.6 ms longer than the average running time of 44.4 ms of the traditional KM algorithm, the improved KM algorithm could achieve the best clustering effect after only one run. Therefore, the improved KM algorithm still achieved a more desirable performance in terms of running time.

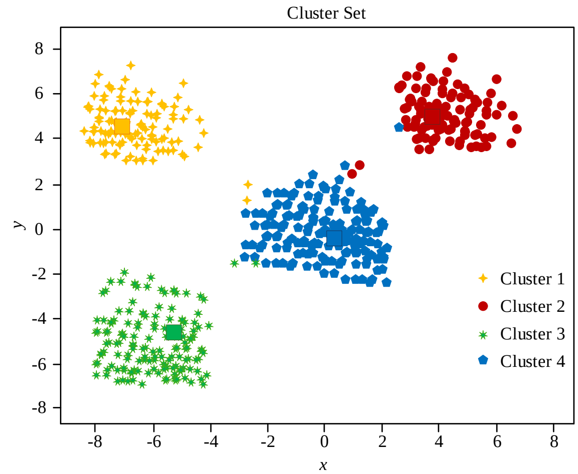
From verifying the traditional and the improved KM algorithms in practical applications, the two algorithms are compared and analyzed in simulation experiments in the Wine dataset. The outcomes of the simulation experimental runs of the two algorithms in the Wine dataset are shown in Fig. 8. Fig. 8(a) indicated that the clustering result of the Wine dataset contained four clusters of almost the same size and no isolated points in the dataset. Fig. 8(b) illustrated that the clustering results of the traditional KM algorithm in the Wine dataset produced misclassification rates of 1.2%, 3.5%, 1.3% and 2.5% for cluster 1, cluster 2, cluster 3 and cluster 4, respectively, and the CA was 97.9%. As seen in Fig. 8(c), the clustering results of the improved KMA in the Wine dataset yielded misclassification rates of 1.3%, 3.5%, 0 and 2.1% for cluster 1, cluster 2, cluster 3 and cluster 4, with a CA of 98.2%.

TABLE II. COMPARISON OF PERFORMANCE INDICATORS BETWEEN TRADITIONAL AND IMPROVED KM ALGORITHMS

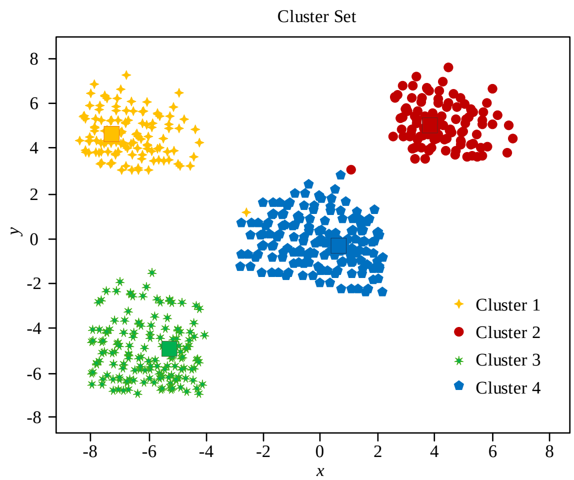
Performance Index	**Data Set**	**Traditional KM Algorithm**			**Improved KM Algorithm**		
		Minimum	***Highest value***	***Average***	***Minimum***	***Highest value***	***Average***
Iterations	Wine	4	12	8.3	7	7	7
	Iris	5	13	9	8	8	8
Clustering accuracy/%	Wine	53.5	70.3	61.9	71	71	71
	Iris	65	90	86.6	90	90	90
Run time/ms	Wine	40	115	75.4	59	59	59
	Iris	26	58	44.4	53	53	53



(a) Clustering results of Wine dataset



(b) Clustering Results of Traditional K-Means Algorithm in Wine Dataset

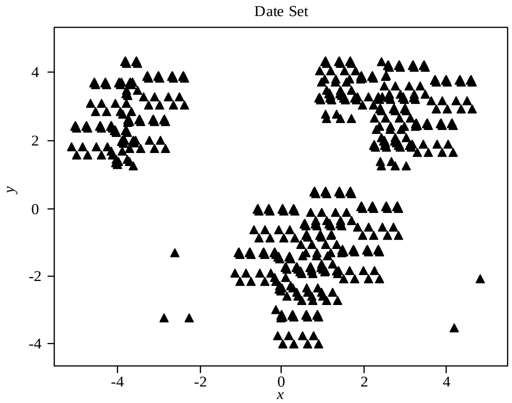


(c) Improved K-Means algorithm for clustering results in Wine dataset

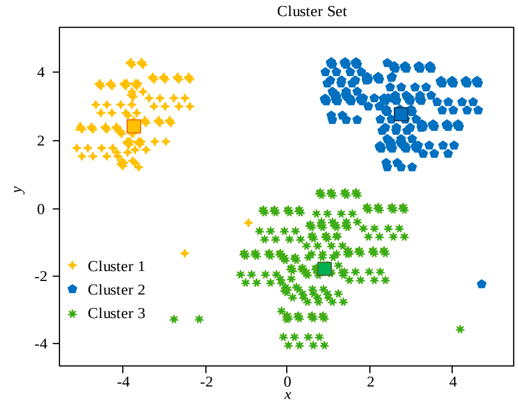
Fig. 8. Simulation experimental results of two algorithms in Wine dataset.

For more comprehensively verifying the traditional and the improved KM algorithms in practical applications, the two algorithms were compared and analyzed in simulation experiments on the Iris dataset. The results of the simulation experimental runs of the two algorithms in the Iris dataset are shown in Fig. 9. Fig. 9(a) demonstrated that the clustering outcome of the Iris dataset contained three clusters of almost the same size and there were five isolated points in the dataset. In Fig. 9(b), the traditional KM algorithm's clustering outcomes on the Iris dataset yielded misclassification rates of 2%, 1.1%, and 4% for clusters 1, 2, and 3, respectively, with a CA of 97.6%. Fig. 9(c) showcased the clustering outcomes of the improved KM algorithm in the Iris dataset, cluster 1, cluster 2, and cluster 3 all produced a misclassification rate of 0, respectively, with a CA of 100%.

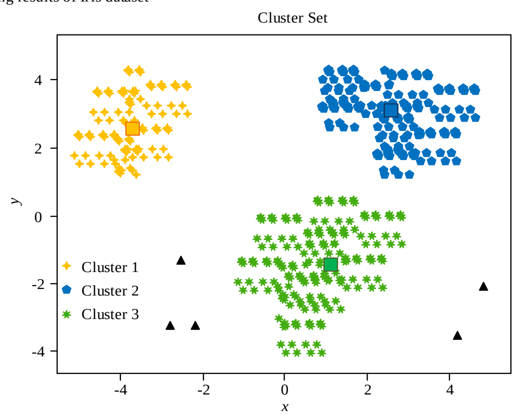
For more intuitively observing the comparison of effectiveness of the traditional and the improved KM algorithms in practical applications, the clustering results in the two datasets were evaluated. The clustering evaluations of the traditional and the improved KM algorithms are shown in Table III. Table III illustrated that the CA of the improved KM algorithm in the Wine dataset was 98.2%, which was 0.3% higher than the CA of 97.9% of the traditional KM algorithm. In the Iris dataset, the CA of the improved KM algorithm reached 100%, which was 2.4% higher than the CA of 97.6% of the traditional KM algorithm. Collectively, the CA of the improved KM algorithm significantly exceeded that of the traditional KM algorithm in practical applications.



(a) Clustering results of Iris dataset



(b) Clustering Results of Traditional KM algorithm in Iris dataset



(c) Improved KM algorithm for clustering results in Iris dataset

Fig. 9. The clustering results of two algorithms in the Iris dataset.

TABLE III. CLUSTER EVALUATION OF TRADITIONAL AND IMPROVED KM ALGORITHMS

Data Set	**Crowd Together**	**Traditional KM Algorithm**			**Improved KM Algorithm**		
		Misclassification rate (%)	***Purity (%)***	***Clustering accuracy***	***Misclassification rate (%)***	***Purity (%)***	***Clustering accuracy***
Wine	Cluster 1	1.2	98.8	97.9	1.3	98.7	98.2
	Cluster 2	3.5	96.5		3.5	96.5	
	Cluster 3	1.3	98.7		0	100	
	Cluster 4	2.5	97.5		2.1	97.9	
Iris	Cluster 1	2	98	97.6	0	100	100
	Cluster 2	1.1	98.9		0	100	
	Cluster 3	4	96		0	100	

V. CONCLUSION

In the context of data ecology, the proliferation of data resources has led to the problems of information pollution and data redundancy in the supply chain service of collection resources. For solving the data redundancy issue and enhancing the collection resource supply efficiency of librarians, the study used DM technology combined with improved KM CA for designing a VCC model of library collection resource supply chain for librarians under the data ecology. The experiment illustrated that the CA of the traditional KM algorithm in the Wine dataset was 70.3% at the highest and 53.5% at the lowest, and the CA of the improved KM algorithm stable at 71%; the CA of the traditional KM algorithm in the Iris dataset was 65% at the lowest and 89% at the highest, and the CA of the improved KM algorithm was maintained at 90%. The running time of the traditional KM algorithm in the Wine dataset was 40ms at the shortest and 115 ms at the longest, and that of the improved KM algorithm is stable at 59 ms; the running time of the traditional KM algorithm in the Iris dataset was 26 ms at the shortest and 58ms at the longest, and that of the improved KM algorithm was maintained at 53 ms. The clustering effect of the improved KM algorithm in practice was 98.2% in the Wine dataset, which was 0.3% higher than the 97.9% CA of the traditional KM algorithm, and 100% in the Iris dataset, which was 2.4% exceeding the 97.6% CA of the traditional KM algorithm. 2.4%. In summary, it illustrates that the studied

VCC model of the supply chain of library resources in view of the data ecology of the library dispensers has better resource supply efficiency, but the experiments use fewer comparison sample data. The experimental results are not objective enough, and further improvement is needed in this aspect. This article studies and explores the influencing factors and mechanisms of VCC in the supply chain of library collection resources from the perspective of data ecology and VCC. Based on the current low utilization rate of library collection resources, differentiated reader needs, and service VCC, improvement strategies are proposed. The conclusions of the study are as follows: firstly, the influencing factors of VCC in the supply chain of library resources mainly include nine major factors: resource construction and integration in the supply chain, development and application of supply chain service technology, data application of supply chain value co creation, staff literacy in the supply chain, concept recognition of supply chain VCC, construction of supply chain VCC platform, subject interaction of supply chain VCC, the reader needs of supply chain VCC and the environmental constraints of supply chain VCC. Secondly, based on the shared supply chain platform and reader decision-making procurement, a VCC model for the collection resource supply chain has been constructed. This model can not only meet the personalized needs of readers, but also improve the utilization rate of collection resources, making the value of the library more fully realized.

REFERENCES

[1] Visser M, Van Eck N J, Waltman L, “Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic,” Quantitative Science Studies, vol. 2, no. 1, pp. 20-41, 2021.

[2] Wu X, Duan J, Pan Y & Li, M, “Medical knowledge graph: Data sources, construction, reasoning, and applications,” Big Data Mining and Analytics, vol. 6, no. 2, pp. 201-217, 2023.

[3] Waziri T A, Ibrahim A, “Discrete Fix Up Limit Model of a Device Unit,” Journal of Computational and Cognitive Engineering, vol. 2, no. 2, pp. 163-167, 2022.

[4] Cui Z, Yan C, “Deep integration of health information service system and data mining analysis technology,” Applied Mathematics and Nonlinear Sciences, vol. 5, no. 2, pp. 443-452, 2020.

[5] Yang Y, Jia F, Xu Z, “Towards an integrated conceptual model of supply chain learning: an extended resource-based view,” Supply Chain Management: An International Journal, vol. 24, no. 2, pp. 189-214, 2019.

[6] Nandi M L, Nandi S, Moya H & Kaynak, H, “Blockchain technology-enabled supply chain systems and supply chain performance: a resource-based view,” Supply Chain Management: An International Journal, vol. 25, no. 6, pp. 841-862, 2020.

[7] Agyabeng-Mensah Y, Ahenkorah E, Afum E, Agyemang, A. N., Agnikpe, C & Rogers, F, “Examining the influence of internal green supply chain practices, green human resource management and supply chain environmental cooperation on firm performance,” Supply Chain Management: An International Journal, vol. 25, no. 5, pp. 585-599, 2020.

[8] Mehrotra S, Rahimian H, Barah M, Luo, F & Schantz, K., “A model of supply-chain decisions for resource sharing with an application to ventilator allocation to combat COVID-19,” Naval Research Logistics (NRL), vol. 67, no. 5, pp. 303-320, 2020.

[9] Zhao Y, Zhang C, Zhang Y, Wang, Z & Li, J, “A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis,” Energy and Built Environment, vol. 1, no. 2, pp. 149-164, 2020.

[10] Liu J, Zhou S, “Application research of data mining technology in personal privacy protection and material data analysis,” Integrated Ferroelectrics, vol. 216, no. 1, pp. 29-42, 2021.

[11] Rong Z, Gang Z. An artificial intelligence data mining technology based evaluation model of education on political and ideological strategy of students. Journal of Intelligent & Fuzzy Systems, vol. 40, no. 2, pp. 3669-3680, 2021.

[12] Wang J, Yang Y, Wang T, Sherratt, R. S & Zhang, J, “Big data service architecture: a survey,” Journal of Internet Technology, vol. 21, no. 2, pp. 393-405, 2020.

[13] Yang J, Li Y, Liu Q, Li, L., Feng, A., Wang, T & Lyu, J, “Brief introduction of medical database and data mining technology in big data era,” Journal of Evidence-Based Medicine, vol. 13, no. 1, pp. 57-69, 2020.

[14] Sutduean J, Singsa A, Sriyakul T & Jermsittiparsert, K, “Supply chain integration, enterprise resource planning, and organizational performance: The enterprise resource planning implementation approach,” Journal of Computational and Theoretical Nanoscience, vol. 16, no. 7, pp. 2975-2981, 2019.

[15] [15] Lou H, “Design of college English process evaluation system based on data mining technology and internet of things,” International Journal of Data Warehousing and Mining (IJDWM), vol. 16, no. 2, pp. 18-33, 2020.

[16] [16] Acquah I S K, Agyabeng-Mensah Y, Afum E, “Examining the link among green human resource management practices, green supply chain management practices and performance,” Benchmarking: An International Journal, vol. 28, no. 1, pp. 267-290, 2020.

[17] John S J, Suma P, Athira T M, “Multiset modules,” Journal of Computational and Cognitive Engineering, vol. 1, no. 1, pp. 37-41, 2022.

[18] Ageed Z S, Zeebaree S R M, 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, vol. 1, no. 2, pp. 29-38, 2021.

[19] Smarandache F, “Plithogeny, plithogenic set, logic, probability and statistics: a short review,” Journal of Computational and Cognitive Engineering, vol. 1, no. 2, pp. 47-50, 2022.

[20] Nandi S, Sarkis J, Hervani A & Helms, M. “Do blockchain and circular economy practices improve post COVID-19 supply chains? A resource-based and resource dependence perspective,” Industrial Management & Data Systems, vol. 121, no. 7, pp. 333-363, 2021.