Intelligent Service Book Sorting in University Libraries Based on Linear Discriminant Analysis Method

Changjun Wang*, Fengxia You, Yu Wang
Library, University of Sanya, Sanya, 572022, China

***Abstract*—The demand for intelligent services in university libraries is constantly increasing, especially in the intelligent book sorting work. The research aims to explore an intelligent classification method for university library books based on linear discriminant analysis. It is used to reduce the dimensionality of feature multidimensional data. A membership model for different categories of books is established to achieve classification. The results showed that when the training set data was reduced to two-dimensional, the feature extraction accuracy of the classification algorithm reached 64.02%, which was significantly higher than 52.48% of one-dimensional data. In addition, the membership calculation accuracy of axiomatic fuzzy sets on two-dimensional data was high, reducing the classification difficulties caused by mixed samples. After comparing and analyzing different algorithms, the proposed transfer learning linear-discriminant analysis-axiomatic fuzzy set algorithm achieved the highest accuracy of 98.67% and completed data classification in about 20s, which was superior to other commonly used classification algorithms. The practical significance of the research lies in providing an efficient and accurate book sorting algorithm, which helps to improve the work efficiency and service quality of libraries.**

***Keywords*—Linear discrimination; library intelligent services; book sorting; university libraries**

I. Introduction

In the digital age, rapidly advancing information technology has not only completely changed learning and working methods, but also reshaped the service model of university libraries [1]. In today's era, university libraries are no longer just storage places for physical books. It has become a digital processing factory for knowledge and an interactive platform for cultural exchange [2]. With the integrated application of big data and intelligent systems, libraries can capture readers' reading habits and preferences in real-time, thereby providing more accurate and personalized services [3]. This transformation has made book management more intelligent, simplifying the process of book classification, retrieval, and borrowing, and greatly improving the user experience [4]. As a key link in intelligent library services, the improvement of automation and intelligence levels in book sorting is particularly important [5]. The current classification of books in libraries mostly relies on manual classification, which not only has low efficiency but also cannot classify books based on their content [6]. Instead, manual classification relies more on book names, which can lead to discrepancies between book categories and actual content. At present, the intelligent book sorting model still has the problem of low classification accuracy in practical applications. The inconsistency between book content and its partition can lead to users' distrust of the library, which is not conducive to the accumulation of users in the library management. Therefore, exploring and applying new intelligent book-sorting methods has become a research hotspot for library service innovation. In order to improve the efficiency and accuracy of book classification in libraries, it is proposed to use linear discriminant analysis technology to mine and classify the feature information of book content. At the same time, a dynamic learning framework was designed to learn the characteristic information of new book types in the library, in order to expand the book classification level of the system.

The study innovatively applies LDA technology to deeply explore the characteristic information of library collections. Based on the basic characteristics of library collections, a user-friendly book sorting assistant tool is designed, which helps to improve the service efficiency of library staff. The main contribution is to develop a standardized book classification evaluation system, which is convenient for other university libraries to reference and learn from, improving the intelligent service level of the entire industry.

The main content of the research is divided into six sections. Section I is an introduction to the significance and purpose of the research. Section II is a survey of related work, which elaborates on the limitations and shortcomings of the current research direction. Section III is a study on the construction of a library book sorting model designed for research. This section is divided into two parts. The first part is the processing method of library book feature information, and the second part is the construction of a book sorting model based on kilometer fuzzy sets. Section IV of the manuscript presents the main results of the research, including training and testing of the constructed model, as well as practical application validation. Section V is a discussion of the research results, comparing the effectiveness of the proposed model with other models in the current field. Section VI is a summary of the main research content.

II. Related Works

LDA is a common statistical analysis method. Ricciardi C et al. proposed a data mining method to analyze the characteristics of patients with myocardial ischemia. 22 features were extracted. The results showed that this method could assist clinical decision-making. The principal component analysis could reduce the number of features [7]. Ontstick L L et al. conducted a double-blind, randomized, controlled clinical trial to identify risk factors associated with mortality in Ebola patients and evaluate the predictive ability of both methods. The results showed that binary logistic regression and LDA selected the same risk factors, predicted similar mortality rates, and scored similar areas under the curve [8]. Bonati L et al. proposed descriptors to characterize these states to obtain appropriate collective variables from metastable information. Neural networks compressed information in low dimensional space and used Fisher LDA as the objective function to maximize the network's discriminative power. This method was tested on alanine dipeptides and nonlinear separable datasets. An inter-molecular aldol condensation reaction was studied [9]. Yan et al. proposed an improved LDA to address the edge classes in multi class classification. The experimental results showed that the self-weighted LDA performed better than other methods, while also having higher computational efficiency [10]. Zhou L et al. conducted a cross-sectional study to investigate the gut microbiota and metabolic function in patients with polycystic ovary syndrome, recruiting participants from different groups. The LDA could effectively count the differences and classify them [11]. Ngailo E K et al. proposed a mis-classification probability approximation algorithm based on the LDA method to solve the mis-classification approximation problem in classification problems. The approximate mis-classification probability of the unknown covariance matrix was derived. The results showed that this method improved the approximation value of mis-classification probability [12].

Intelligent services in university libraries are a necessary condition for the development of university libraries in the information age. Mohammed SH et al. used topic modeling method to solve the classification problem of Chinese text data in digital libraries. Comparing the two popular methods of Latent Dirichlet Allocation and latent semantic allocation, the former performed better. When the topics were 20, the highest coherence of the potential Dirichlet allocation was 0.592179, while the potential semantic allocation was 0.577302 [13]. Asemi A et al. used descriptive and content review methods to review articles related to robotics in libraries and information science from 2007 to 2017. The intelligent systems contributed to multiple purposes for librarians, including technical services, public services, and consultation desks. Through integration with artificial intelligence technology, current information systems had high potential for improvement [14]. Wan M et al. proposed a cloud-based product service system platform to address the challenge of urban waste issues. This platform aimed to manage solid waste management resources. Case analysis verified the feasibility of the platform [15]. Chiu CK et al. proposed a learning status management system in an intelligent classroom to improve the learning status of students. Sensor technology and image recognition technology were used to detect and collect various information from students, and Bayesian classification networks were used to infer the learning status. In addition, the system also included a feedback mechanism. Two experiments verified the accuracy and effectiveness. The learning status analysis was highly consistent with observed results. Students were more focused in class [16].

The current research on book sorting still adopts the method of classifying books based on their semantic names, which leads to significant deviations in the accuracy of book collection and sorting. Moreover, the current intelligent management platform for library management still faces problems of inability to retrieve or inaccurate retrieval results in practical applications. Aiming at the shortcomings of the current intelligent management and sorting platform for library management, this study proposes the use of LDA technology to extract feature information of key content of books and sort them based on the key content features of books. The research has changed the main sorting basis of current book sorting technology, adopting a new sorting mode to achieve the goal of improving the accuracy of book sorting.

In summary, LDA is a commonly used statistical method that has been frequently applied in various studies. Influenced by information technology, university libraries are no longer limited to collecting images and are beginning to focus on intelligent service systems. Classification management technology has rich applications in various fields. To improve the service quality, the intelligent service system of university libraries is utilized to achieve the classification management of books. Therefore, the study combines LDA with intelligent services in university libraries to conduct book-sorting research.

III. Book Sorting Model based on LDA and AFS

With the development of the information age, university libraries have also begun to develop intelligent services. Intelligent book sorting can effectively reduce library management costs and improve library classification accuracy. In Section II, LDA and AFS are used to study the book sorting methods in university libraries. A book sorting model is constructed. The research will be conducted from two aspects. The first is data dimensionality reduction based on linear discrimination, and the second is the image book sorting model based on AFS.

A. Data Dimensionality Reduction Based on Linear Discriminant

LDA is a classic statistical method widely applied in pattern recognition and machine learning. It is a supervised learning algorithm mainly used for dimensionality reduction and classification problems [17]. The core idea is to project high-dimensional data onto a low-dimensional space that best distinguishes categories. Specifically, LDA attempts to find a series of projection axes such that when data is projected onto these axes. The distance between different categories is as large as possible, and the data dispersion within the same category is as small as possible, as shown in Fig. 1 [18].

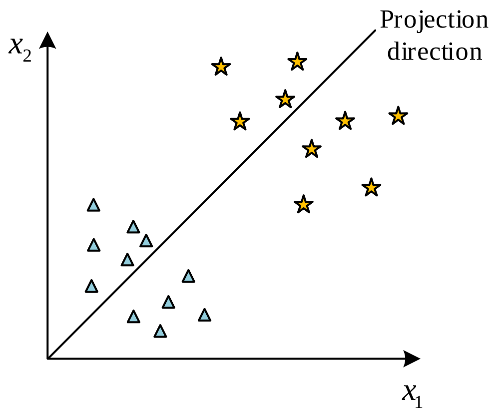


Fig. 1. LDA.

In the LDA algorithm, it is assumed that the data sample set is D. This sample set contains multiple data points. Each data point x_i can be considered as a vector in an n-dimensional space. Each component of this vector represents a feature of the data. Each data point x_i usually has an associated label category y_i, which indicates the classification to which the data point belongs. These categories are crucial for establishing classification models and making predictions in the future. In each category, it is assumed to be a classification labeled c. These classifications all have a corresponding sample size N_c. The number of samples determines the representativeness of each classification in the sample set. Generally speaking, a larger sample size indicates that the statistical characteristics of the corresponding classification are more reliable. To further analyze each category, the set of sample points for each data category is defined as X_c. For each category c, two important statistical measures can be further defined: the mean vector μ_c and the covariance matrix Σ_c. The mean vector is shown in Eq. (1) [19].

$$\mu_c = \frac{1}{N}\sum_{x \in X_c} x, c = (0,1) \tag{1}$$

The covariance matrix is shown in Eq. (2).

$$\Sigma = \sum_{x \in X_c} (x-\mu_c)(x-\mu_c)^T, c = (0,1) \tag{2}$$

A common method when dealing with binary classification problems is to find an appropriate mapping that reduces the dimensionality of high-dimensional data sample points to one-dimensional space to simplify the classification process. In mathematics, this can be achieved by defining a normal vector w, which essentially describes a vector in a one-dimensional spatial direction. Through this normal vector, the projection of each data sample point x_i on the line is represented as $w^T x_i$. Similarly, the mean values of different categories of sample points in the dataset can also be projected onto this vector to generate new mean points $w^T\mu_0$ and $w^T\mu_1$. Based on these criteria, LDA finds the optimal projection direction w through a mathematical optimization process. The original binary classification problem is transformed into a simple one-dimensional decision problem. In this way, no matter how high the dimensions of the original data are, theoretically, it can be effectively classified. The optimization objective is shown in Eq. (3).

$$MaxJ(w) = \frac{\left\| w^T\mu_0 - w^T\mu_1 \right\|_2^2}{w^T \sum_0 w + w^T \sum_1 w} = \frac{w^T(\mu_0-\mu_1)(\mu_0-\mu_1)^T w}{w^T\left(\sum_0 w + w \sum_1 w\right) w} \tag{3}$$

In Eq. (3), $MaxJ(w)$ refers to the distance from the sample point to the center point. If $(\mu_0-\mu_1)(\mu_0-\mu_1)^T$ is defined as s_b and $\sum_0 w + w \sum_1 w$ is defined as s_w, then it can be expressed as Eq. (4).

$$s_w = \sum_{x \in X_c} (x-\mu_0)(x-\mu_0)^T + \sum_{x \in X_c} (x-\mu_1)(x-\mu_1)^T \tag{4}$$

By combining Eq. (4) with Eq. (3) and simplifying them, Eq. (5) can be obtained.

$$MaxJ(w) = \frac{w^T s_b w}{w^T s_w w} \tag{5}$$

Eq. (5) is the generalized Rayleigh quotient. In practical applications, most data samples are two or more categories. In this case, the optimization objective of LDA sample classification can be rewritten as Eq. (6).

$$\frac{W^T S_b W}{W^T S_w W} \tag{6}$$

In Eq. (6), S_b refers to the inter class discrete matrix of the sample points. S_w refers to the intra class discrete matrix of the data sample. S_b is shown in Eq. (7).

$$S_b = \sum_{c=1}^{k} N_c(\mu_c-\mu)(\mu_c-\mu)^T \tag{7}$$

S_w is displayed in Eq. (8).

$$S_w = \sum_{c=1}^{k}\sum_{x \in X_c}(x-\mu_c)(x-\mu_c)^T \tag{8}$$

From Eq. (7) and Eq. (8), the optimization objectives of two-class LDA cannot optimize multi-class LDA objectives. Therefore, other optimization methods are adopted to optimize multi-class LDA. The optimization objective function of multi-class LDA can be defined, as displayed in Eq. (9).

$$MaxJ(W) = \frac{\prod_{diag} W^T S_b W}{\prod_{diag} W^T S_b W} \tag{9}$$

In Eq. (9), $\prod_{diag} W^T S_b W$ refers to the diagonal element product. W represents a matrix. The optimization model for multi-class LDA can be transformed into Eq. (10).

$$MaxJ(W)=\prod_{i-1}^{d}\frac{W_i^T S_b W_i}{W_i^T S_b W_i} \tag{10}$$

LDA is a common data dimensionality reduction method. When LDA performs data dimensionality reduction, it needs to calculate the intra class and inter class divergence matrix, and matrix $S_W^{-1}S_b$. When calculating $S_W^{-1}S_b$, a Lagrangian function needs to be introduced. The maximum eigenvalue $S_W^{-1}S_b$ and corresponding eigenvector w can be calculated. Fig. 2 displays the specific process [20].

Manifold Learning is a type of machine learning method used for dimensionality reduction, aimed at discovering and utilizing the underlying structure of the data itself to reveal its intrinsic features. These methods are based on the manifold assumption, which means that high-dimensional data is actually distributed on an unknown low-dimensional manifold. This low-dimensional manifold is implant in a high-dimensional space. The implementation of manifold learning is shown in Fig. 3 [21].

The manifold learning is applied to mine the potential structures of data in high-dimensional data spaces and obtain low-dimensional manifold representations. Then, through LDA, the low-dimensional manifold is mapped to a more discriminative space. The process of constructed Transfer Learning-LDA (TL-LDA) algorithm is shown in Fig. 4.

B. Book Sorting based on AFS

A fuzzy set is a mathematical theory used to handle uncertainty and fuzzy information, which can effectively deal with fuzzy and imprecise problems. The Membership Function (MF) of fuzzy sets can be used to describe the characteristics and attributes of fuzzy sets. AFS is a form of fuzzy set theory that satisfies a specific set of axioms and operational rules. These axioms and rules define the basic properties and operation methods of fuzzy sets, making the operations and reasoning of fuzzy sets consistent and reliable. AFS can model, represent, and process fuzzy concepts to solve real-world fuzzy problems. The research content of AFS includes membership functions of fuzzy sets, operation rules, feature descriptions, and inference methods, as shown in Fig. 5.

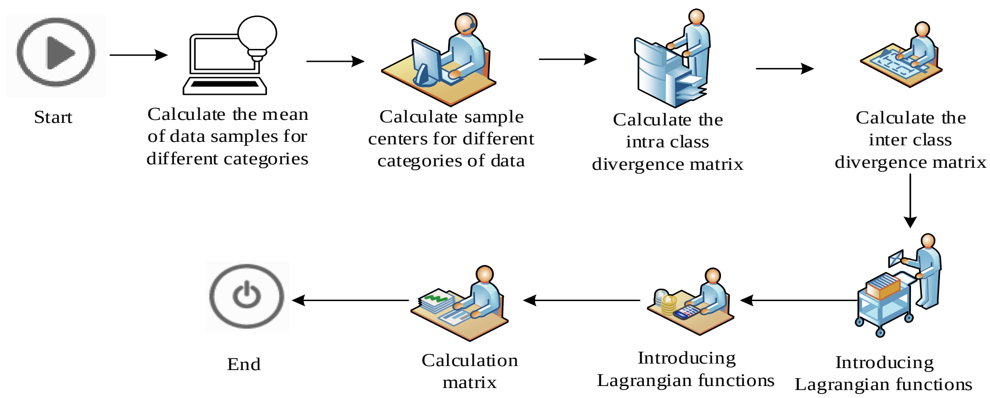


Fig. 2. LDA flow path.

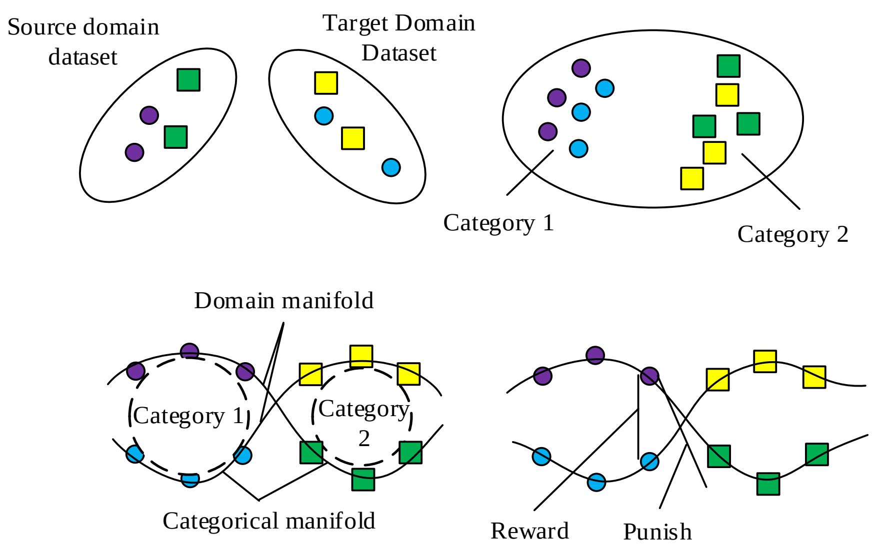


Fig. 3. The manifold learning framework.

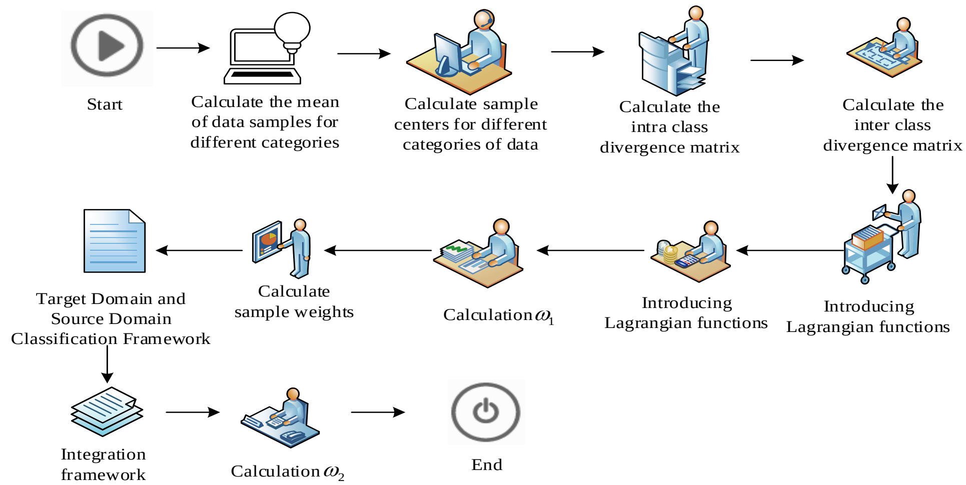


Fig. 4. Transfer learning - LDA algorithm process.

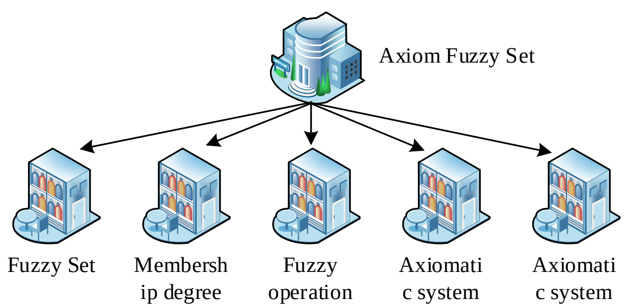


Fig. 5. Axiomatic fuzzy set of basic concepts.

Fuzzy sets refer to traditional set theory where an element either belongs to a set or not. In fuzzy set theory, the MF concept is introduced to represent the degree to which an element belongs to a certain set, which is represented by a numerical value in [0, 1]. Membership degree is the degree to which an element corresponds to a fuzzy set, reflecting the agreement between that element and the features of the fuzzy set. The MF is applied to determine the membership degree of an element in a fuzzy set. Fuzzy operations include fuzzy intersection, union and complement, which correspond to the traditional set theory. However, in their definition and operation process, the membership degree of the elements is considered. Axiom systems are designed to construct a robust fuzzy logic system. The AFS method is based on a set of strictly defined axioms that describe the basic rules that membership functions should follow, ensuring theoretical consistency and closure. Approximation reasoning refers to the reasoning process in fuzzy systems based on known information and axiomatic rules for unknown or uncertain situations. If M is assumed to be a simple set of concepts on Ω, (Ω, F, P) represents the probability measure space, and the observation dataset of (Ω, F, P) is represented as $X \subset \Omega$, (M, τ, Ω), and $(M, \tau | x, X)$. ξ is a concept that conforms to Eq. (11).

$$\xi = \sum_{i \in I} \left(\prod_{m \in A} m \right) \in EM \tag{11}$$

In Eq. (11), M represents a simple concept, which belongs to the set A. The membership function of AFS can be defined as Eq. (12).

$$\mu\xi(x) = \sup_{i \in I} \inf_{\gamma \in A_i} \frac{\sum_{u \in A_i^\tau} \rho_\gamma(u) N_u}{\sum_{u \in X} \rho_\gamma(u) N_u}, \forall x \in X \tag{12}$$

In Eq. (12), N_u represents the number of times sample $x \in X$ has been observed in the dataset. ρ_γ represents the weight function of a simple concept. $\rho_\gamma(x)$ is continuous on $\gamma(x)$. X is a set of randomly obtained data in probability space, and $\gamma \in M$. Then for $\forall x \in \Omega$, when $|X|$ approaches infinity, Eq. (12) can converge to Eq. (13).

$$\mu\xi(x) = \sup_{i \in I} \inf_{\gamma \in A_i} \frac{\int_{A_i^\tau(x)} \rho_\gamma(t) dP(t)}{\int_{\Omega} \rho_\gamma(t) dP(t)}, \forall x \in \Omega \tag{13}$$

After determining the membership function of the AFS structure, a book sorting model is constructed. The model uses TL-LDA to reduce the dimensionality of book data, extract the corresponding features of these data, and calculate the weight ratio of each data. AFS calculates the eligibility of each data, obtains a brief explanation, and then completes the corresponding data classification based on the brief explanation. The specific process is shown in Fig. 6.

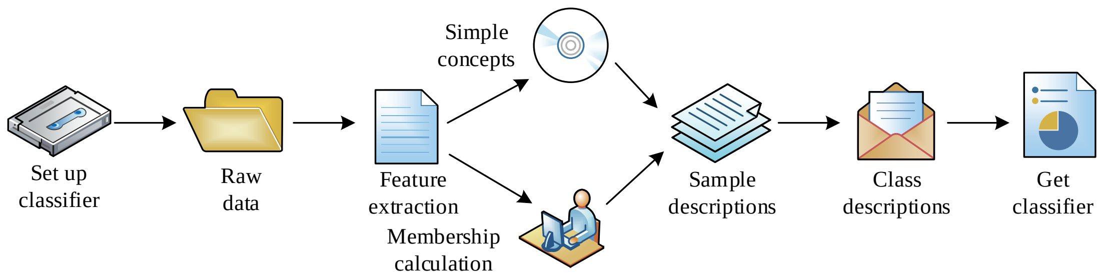


Fig. 6. Process of book sorting algorithm based on TL-LDA-AFS.

IV. Simulation Experimental Analysis of Book Sorting Model based on LDA and AFS

In Section II, a book sorting model based on LDA and AFS is constructed. To verify the effectiveness, simulation experiments and analysis are designed on the model in Section III. Section III is divided into two parts. The first part sets the experimental environment and parameters, and the second part analyzes the simulation experiment results.

A. Experimental Environment and Parameter Settings

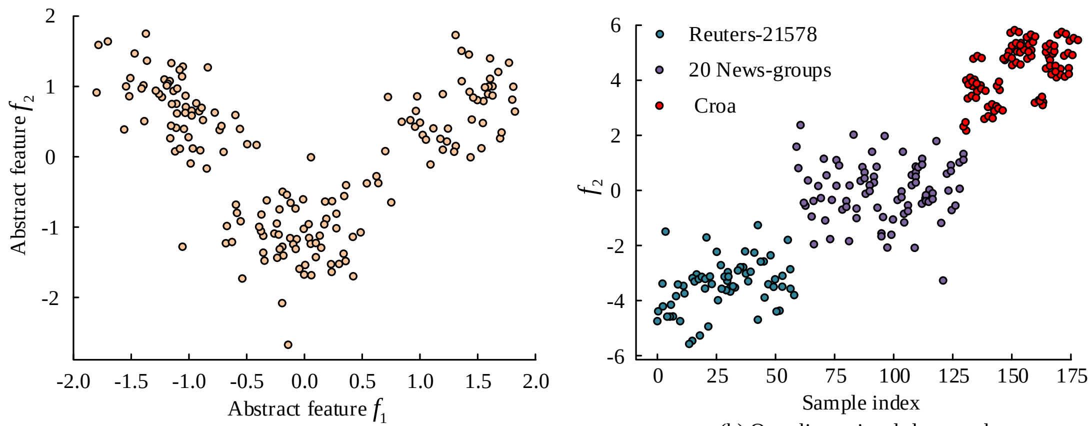
The experimental environment for the study is built in Python. The operating system is Windows 7 64bit, the processor is Inter (R) Core i5-12440, and the memory size is 16GB. The dataset used in the study includes the Reuters-21578 dataset, the 20 News-groups, and the Cora. Reuters-21578 is an iconic dataset in text classification, consisting of 21578 documents that provide detailed records of various major events and industry dynamics that have occurred globally. 20 The 20 News-groups is a large-scale text dataset designed specifically for research in text mining and text classification. The Cora dataset is an online document library containing approximately 37000 detailed computer science research papers in computer science research. These three datasets are used to generate 12 pairs of datasets as training and detection datasets for the model. The data ratio between the training set and the test set is 7:1, as displayed in Table I.

B. Data Dimensionality Reduction Processing Analysis

The TL-LDA performs LDA to reduce the dimensionality of the training set data. The results are shown in Fig. 7.

TABLE I. Experimental Dataset

Data number	Training task		Detection task	
	Positive	Negative	Positive	Negative
1	Orgs (1)	People (1)	Orgs (2)	People (2)
2	Place (1)	poisonous(enlarging)	Place (2)	Poisonous (tapering)
3	People (1)	edible(enlarging)	People (2)	edible(tapering)
4	encryption	protocols	compression	routing
5	protocols	Probabilistic methods	routing	genetic algorithms
6	encryption	probabilistic methods	compression	algorithms
7	comp.os	sci.crypt	comp.mac	sci.space
8	sci.med	rec.sport.baseball	talk.politics.gun	rec.sport.hokey
9	talk.religion	talk.religion	sci.crypt	talk.politics.guns
10	Computation-complexity	encryption	Computation-geometry	compression
11	Computation-complexity	probabilistic methods	Computation-geometry	genetic algorithms
12	Computation-complexity	protocols	Computation-geometry	routing



(a) LDA dimensionality reduction analysis

(b) One-dimensional data results

Fig. 7. LDA dimension reduction.

Fig. 7(a) shows the results of LDA dimensionality reduction on the training set data. The main data was concentrated in three regions. Fig. 7(b) shows the results of three datasets after reducing the training dataset to 1 dimension. After the data is reduced to 1 dimension, the three basic data exhibited a linear distribution. Samples of different categories may be confused within certain value ranges. The membership degree of these confused samples in specific environments is analyzed in detail to determine the MF. After the data is reduced to 1 dimension, the number of data features decreases, making it difficult for the model to distinguish these confusing samples. Therefore, reducing the dimensionality to 2D data is sufficient when performing dimensionality reduction processing. In situations with more features, more features can be applied to create fuzzy sets, which enhances classification ability. When performing discrimination, AFS can make judgments based on the membership degree on different concepts, further improving the universality and accuracy. However, it's not that having more features leads to better performance. Excessive features may lead to the curse of dimensionality in fuzzy set rules. In addition to the difficulty of model parsing, this may also consume more computing resources. Therefore, balancing the number of features is crucial for building an effective fuzzy set model. To further determine the degree of pre-processing, the accuracy between one-dimensional data and two-dimensional data is compared. The results are shown in Fig. 8.

In Fig. 8, regardless of whether the data was reduced to one-dimensional or two-dimensional, the accuracy in extracting data features continuously decreased with the increase of the ratio. When the ratio was 0.8, the accuracy of feature extraction for both dimensions decreased to the lowest. The accuracy of feature extraction for one-dimensional data was only 0.3248, while the two-dimensional data was 0.4676, which was 0.1428 higher than the one-dimensional data. When the ratio was 0.1, the accuracy of feature extraction in both dimensions increased to the highest. The accuracy of feature extraction for one-dimensional data was 0.5248, and the two-dimensional data was 0.6402, which was 0.1154 higher than that of one-dimensional data. The feature extraction accuracy of two-dimensional data was equal before the ratio was less than 0.4. The accuracy of feature extraction in one-dimensional data was always changing. When the ratio was 0.5, the accuracy of feature extraction in two-dimensional data was 0.5964, while the one-dimensional data was only 0.4856, which was 0.1008 lower than that in two-dimensional data.

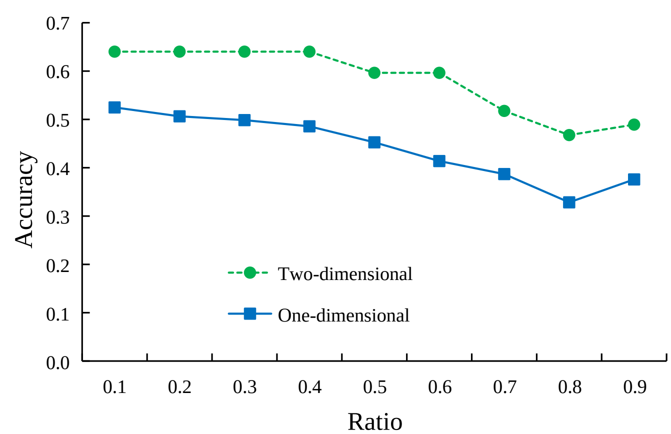


Fig. 8. Comparison of the feature extraction accuracy rate.

C. Analysis of Classification Results

After the dataset is dimensionally reduced by LDA, the membership degree is calculated using the reduced data. The results are shown in Fig. 9.

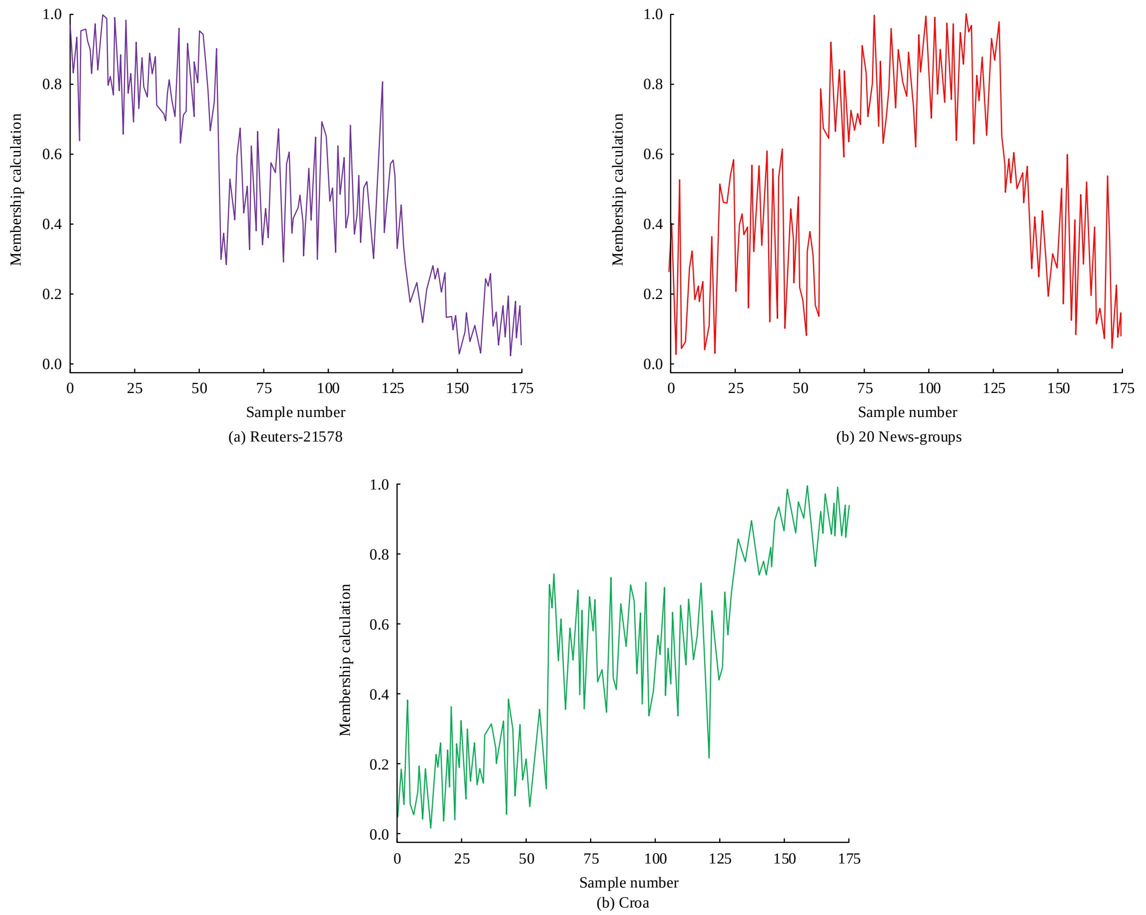


Fig. 9. Membership levels on the different datasets.

Fig. 9(a) shows the membership degree on the Reuters-21578 dataset. The membership degree of the top 50 samples was concentrated between 0.6 and 1.0. The membership degree of samples 50 to 125 was concentrated between 0.2 and 0.6. Samples 125 to 175 were concentrated between 0.0 and 0.3. Fig. 9(b) shows the membership degree on the 20 News-groups dataset. The membership degree of the top 50 samples was concentrated between 0.0 and 0.6. Samples 50 to 125 were concentrated between 0.6 and 1.0. Samples 125 to 175 were concentrated between 0.0 and 0.6. Fig. 9(c) shows the membership degree on the Cora dataset. The membership degree of the top 50 samples was concentrated between 0.0 and 0.4. Samples 50 to 125 were concentrated between 0.2 and 0.8. Samples 125 to 175 were concentrated between 0.8-1.0. The study describes the membership degrees of three datasets, as shown in Fig. 10.

Fig. 10 shows the membership degree distribution of three categories relative to all samples in the same chart. It is possible to visually observe the membership values of each category on different samples. The significant differences in membership indicate that classification decisions can be made based on these differences, thereby improving the accuracy and efficiency of classification. There are significant differences in the membership degrees of the three different datasets across all samples. Support Vector Machine (SVM), K-nearest Neighbors (KNN), Native Bayes (NB), and Logistic algorithms are commonly used classification algorithms. To verify the feasibility of TL-LDA-AFS, the accuracy and classification time of these algorithms are compared on the dataset. The results are shown in Fig. 11.

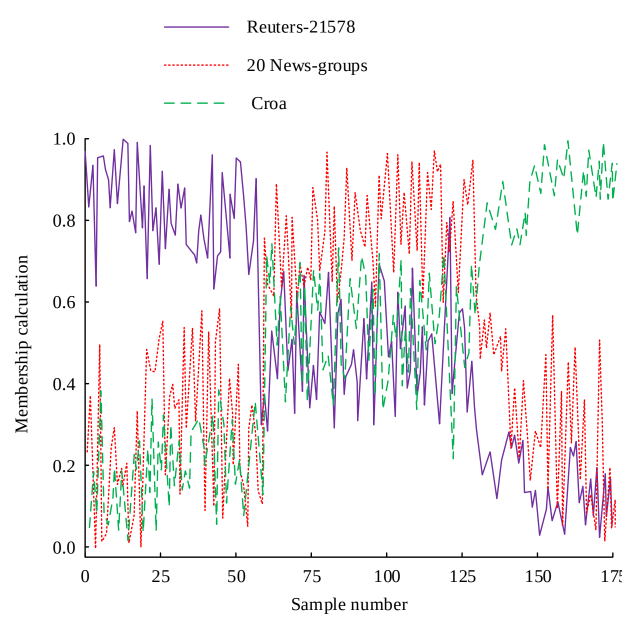
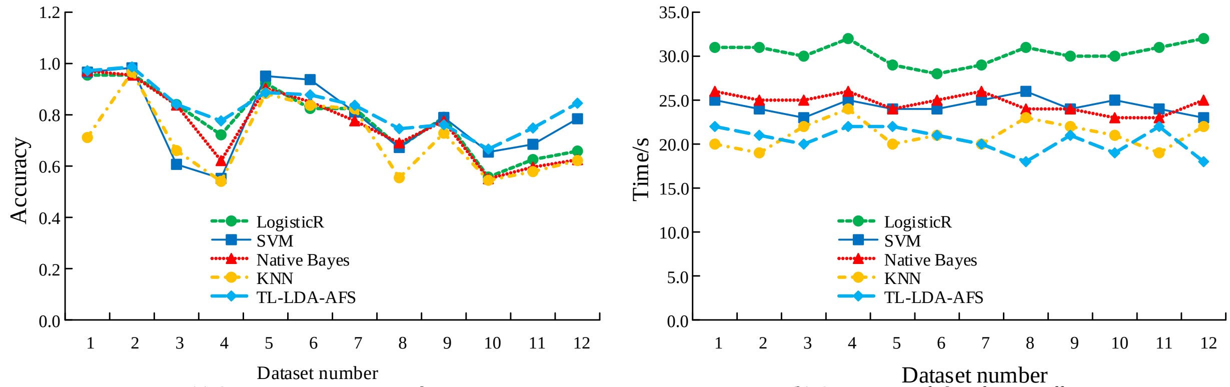


Fig. 10. Class description membership comparison.



(a) Accuracy comparison results
(b) Comparison of classification efficiency

Fig. 11. Model feasibility validation.

Fig. 11(a) shows the accuracy comparison. In dataset 2, all algorithms achieved the highest accuracy. The accuracy of SVM, KNN, NB, Logistic, and TL-LDA-AFS algorithms were 0.9833, 0.9660, 0.9547, 0.9600, and 0.9867, respectively. In dataset 10, all algorithms had the lowest accuracy. The accuracy of SVM, KNN, NB, Logistic, and TL-LDA-AFS algorithms were 0.6545, 0.5445, 0.5507, 0.5572, and 0.6662, respectively. Fig. 11(b) shows the algorithm time consumption. On all datasets, the TL-LDA-AFS algorithm had the lowest classification time, taking about 20s to complete the data classification. SVM, KNN, NB, and Logistic took 24s, 20 seconds, 25s, and 30s to complete data classification, respectively. Whether it is the accuracy of data classification or classification efficiency, the TL-LDA-AFS algorithm has higher accuracy. In text datasets, its performance is significantly better than other classification algorithms.

V. Discussion

A study has proposed an intelligent book classification method based on LDA and AFS to improve the efficiency and accuracy of book classification in university libraries. Asemi et al. [14] found through content review methods that intelligent systems contribute to multiple aspects of libraries. Therefore, the study adopts artificial intelligence technology to design a book sorting model. Mohammed et al. [13] used topic modeling methods to solve the classification problem of Chinese text data in digital libraries and found that latent Dirichlet allocation performed better than latent semantic allocation. Similar to the research findings, the use of artificial intelligence technology can effectively improve the level of book sorting.

The experimental results show that when the data is reduced to two-dimensional, the accuracy of feature extraction reaches 64.02%, which is significantly higher than the 52.48% of one-dimensional data. The transfer learning linear discriminant analysis axiomatic fuzzy set (TL-LDA-AFS) algorithm proposed in the study achieved the highest classification accuracy of 98.67% among all comparison algorithms, and completed data classification in approximately 20 seconds, significantly better than other commonly used classification algorithms. The intelligent book classification method based on LDA and AFS proposed in the study not only improves the accuracy and efficiency of book classification, but also provides a new intelligent solution for book management in university libraries.

VI. Conclusion

The research aims to create an intelligent book sorting system for university libraries to improve book management efficiency and user experience. Based on the LDA method, this study successfully reduces the dimensionality of multidimensional feature data to two-dimensional, making it easier to classify books more accurately. From the experimental results, when the data was reduced to 1 dimension, the three basic data exhibited a linear distribution. When the ratio was 0.8, the accuracy of feature extraction for both dimensions decreased to the lowest. The accuracy of feature extraction for one-dimensional data was only 0.3248, while the accuracy for two-dimensional data was 0.4676, which was 0.1428 higher than the one-dimensional data. The accuracy of feature extraction for LDA processed data in two-dimensional dimensionality reduction was 64.02%, which was approximately 11.54% higher than that of one-dimensional data. After comparing different classification algorithms, the TL-LDA-AFS algorithm developed in this study achieved the highest accuracy of 98.67% on dataset 2, which was the highest among all comparison algorithms. It also showed significant advantages in classification efficiency, taking only about 20s to complete classification, greatly improving the intelligence level of book sorting in university libraries. The study provides a fast and accurate intelligent book sorting algorithm for university libraries. It is expected to promote the progress of library management towards intelligence and automation, while improving the quality and efficiency of library services for teachers and students. The book sorting model designed for research has a large amount of data computation and high system load requirements when learning the feature information of different categories of books. It takes a long time to process the newly added book categories in the library, resulting in a high resource load on the book sorting system. Future research will further optimize the model's ability to learn feature information for new types of books, reduce the model's learning time for new classifications, and lower the model's burden on system resources.

REFERENCES

[1] F. Zhang, "Research on Keyword Mining of Academic Library Subject Service," Acta Informatica Malaysia (AIM), vol. 7(2), pp. 97-100, 2023.

[2] C. Daraojimba, O. Onunka, T. Onunka, A. A. Fawole, I. J. Adeleke, "Library and Information Services in The Digital Age: Opportunities and Challenges," Acta Informatica Malaysia (AIM), vol. 7(2), pp. 113-121, 2023.

[3] Lund B D, Wang T. Chatting about ChatGPT: how may AI and GPT impact academia and libraries. Library Hi Tech News, 2023, 40(3): 26-29.

[4] Winata A P, Fadelina R, Basuki S. New normal and library services in Indonesia: A case study of university libraries. Digital Library Perspectives, 2021, 37(1): 77-84.

[5] Lee P C. Technological innovation in libraries. Library Hi Tech, 2021, 39(2): 574-601.

[6] Tella A. Robots are coming to the libraries: are librarians ready to accommodate them. Library Hi Tech News, 2020, 37(8): 13-17.

[7] Ricciardi C, Valente A S, Edmund K, Cantoni V, Green R, Fiorillo A, Cesarelli M. Linear discriminant analysis and principal component analysis to predict coronary artery disease. Health Informatics Journal, 2020, 26(3): 2181-2192.

[8] Ontshick L L, Sabue J C M, Kiangebeni P M, Mbayi O T, Ntamabyaliro J M N, Tamfumu J J M, Matendo R M M. Comparison of the performance of linear discriminant analysis and binary logistic regression applied to risk factors for mortality in Ebola virus disease patients. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 2023, 5(3): 205-210.

[9] Hasanvand M, Nooshyar M, Moharamkhani E, Selyari A. Machine Learning Methodology for Identifying Vehicles Using Image Processing. AIA, 2023, 1(3): 170-178.

[10] Yan C, Chang X, Luo M, Zheng Q, Zhang X, Li Z, Nie F. Self-weighted robust LDA for multiclass classification with edge classes. ACM Transactions on Intelligent Systems and Technology (TIST), 2020, 12(1): 1-19.

[11] Zhou L, Ni Z, Cheng W, Yu J, Sun S, Zhai D, Cai Z. Characteristic gut microbiota and predicted metabolic functions in women with PCOS. Endocrine Connections, 2020, 9(1): 63-73.

[12] Ngailo E K, Chuma F. Approximation of misclassification probabilities in linear discriminant analysis based on repeated measurements. Communications in Statistics-Theory and Methods, 2023, 52(23): 8388-8407.

[13] Mohammed S H, Al-augby S. Lsa & lda topic modeling classification: Comparison study on e-books. Indonesian Journal of Electrical Engineering and Computer Science, 2020, 19(1): 353-362.

[14] Asemi A, Ko A, Nowkarizi M. Intelligent libraries: a review on expert systems, artificial intelligence, and robot. Library Hi Tech, 2020, 39(2): 412-434.

[15] Wan M, Qu T, Huang M, Li L, Huang G Q. Cloud-based product-service systems platform for household solid waste classification management. IET Collaborative Intelligent Manufacturing, 2020, 2(2): 66-73.

[16] Chiu C K, Tseng J C R. A bayesian classification network-based learning status management system in an intelligent classroom. Educational Technology & Society, 2021, 24(3): 256-267.

[17] Kim S, Rigatto K, Gazzana M B, Knorst M M, Richards E M, Pepine C J, Raizada M K. Altered gut microbiome profile in patients with pulmonary arterial hypertension. Hypertension, 2020, 75(4): 1063-1071.

[18] Xu C, Shen J. Multi-criteria decision making and pattern recognition based on similarity measures for Fermatean fuzzy sets. Journal of Intelligent & Fuzzy Systems, 2021, 41(6): 5847-5863.

[19] Bhattacharyya R, Mukherjee S. Fuzzy membership function evaluation by non-linear regression: An algorithmic approach. Fuzzy Information and Engineering, 2020, 12(4): 412-434.

[20] Lazaroiu G C, Roscia M. Fuzzy logic strategy for priority control of electric vehicle charging. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 19236-19245.

[21] Jan N, Gwak J, Pei J, Maqsood R, Nasir A. Analysis of networks and digital systems by using the novel technique based on complex fuzzy soft information. IEEE Transactions on Consumer Electronics, 2022, 69(2): 183-193.