

Machine Learning in the Construction of Library's Special Collection Document Information Resources

Jie Ren*

College of Physical Education and Training, Harbin Sports University, Harbin 150008, China

renjie@hrbipe.edu.cn

**corresponding author*

Keywords: Machine Learning, Library Special Collections, Information Resources, Sharing Construction

Abstract: With the rapid development of Internet technology, more and more people begin to focus on the construction of library (CL) information resources (IR) in the field of mobile network and social media applications. In order not to let the unique culture disappear gradually, and to realize cultural exchange and inheritance, this paper intends to use modern technology to build IR of library related documents. This paper mainly uses the principal component analysis and survey methods to carry out relevant research on the CL special collection document IR. The survey results show that 70% of people think that the utilization rate of library resources is not high, while 68% of people think that the research on special collection culture should be strengthened.

1. Introduction

With the development of Internet technology, information resource sharing has become a new trend. The library is the most important hub for the cultural and economic exchange records of various places. This paper aims to study how to build efficient, stable and high-quality shared digital content, and use this system for user analysis and retrieval, classification and recommendation related knowledge query. Therefore, it is of practical significance to study the construction model of library special collection document system.

There are many research theories on the CL special collection document IR and online resource sharing. For example, some scholars pointed out that more than 500000 kinds of books are published every year in China, which makes it difficult to find and collect special collection literature resources in libraries [1-2]. Some scholars believe that the recommendation system based on artificial intelligence can effectively alleviate the information overload and solve the practical problems of special collection literature [3-4]. In addition, some scholars elaborated on the significance of the multimedia of library characteristic collection resources, the current situation of

the construction and application of multimedia resources, and the impact of 5G era on the multimedia of characteristic collection [5-6]. Therefore, the research on the CL special collection document IR in this paper is of great significance and social value.

This paper first studies the construction of digital resources of local documents under the network environment. Secondly, it expounds the problems existing in the construction and utilization of library IR. Then it puts forward the countermeasures for the CL IR. Then it briefly describes the application of machine learning in the CL resources. Finally, through the form of questionnaire, ask others for suggestions, and get relevant solutions.

2. Machine Learning and the Construction of Library Literature Information Resources

2.1. Construction of Local Document Digital Resources under Network Environment

The wide application of multimedia resources has strengthened the information awareness of local document resource interviewers. The digitalization of document carriers has enriched the forms of document resources in libraries, including electronic magazines, audio and video CDs, electronic courseware, etc. With the realization of network services, more readers can vividly display multimedia literature content by means of network or electronic borrowing. In this paper, machine learning can be used to classify and sort out the literature information in the library. The common classification methods include principal component analysis [7-8]. The original PCA dimension reduction steps are shown in Figure 1:

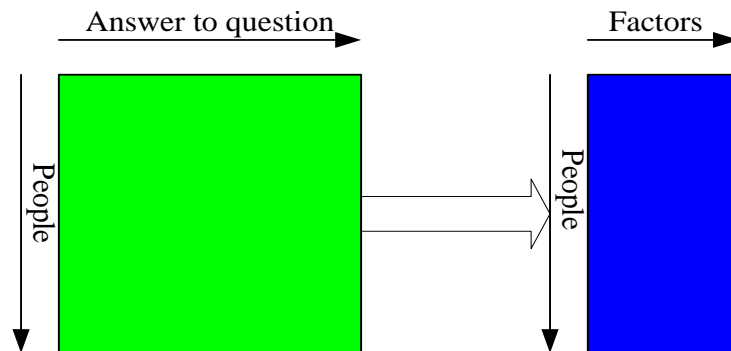


Figure 1. Start the PCA dimension reduction step

First, calculate the covariance matrix of the original data. The solution of the eigenvalue of the matrix is $\mu_1 \geq \mu_2 \geq \dots \geq \mu_q \geq 0$. Finally, calculate the variance contribution rate of each principal component:

$$\tau_m = \frac{\mu_m}{\sum_{k=1}^q \mu_k}, m = 1, 2, \dots, q \quad (1)$$

The variance contribution rate of each principal component reflects the amount of original information contained in the component. Finally, calculate the variance cumulative variance contribution rate of the first X principal components:

$$\tau_x = \frac{\sum_{m=1}^x \mu_m}{\sum_{k=1}^q \mu_k}, m \pi q \quad (2)$$

When the cumulative variance contribution rate reaches more than 85%, we think it is ideal.

In the network environment, the construction of local literature IR has cost and expense considerations. For example, the substantial increase in collection costs, the scattered and disordered collection of literature, the embedded use of computer network technology and communication technology, etc.. All of these objectively forced local libraries to strengthen mutual cooperation and coordination in the construction and accumulation of local literature IR. In the construction of local literature resources under the network environment, we should first digitize the existing local literature resources, form our own information products, and increase the online information stock. Resource exchange and sharing is very intensive and economic from the perspective of development, which accumulates resources and avoids repeated development. Reference to the resource exchange and sharing of local cultural institutions means that all participating museums can share other data resources while providing their own collection data [9-10].

2.2. Problems in the Construction and Utilization of Library IR

At present, the degree of book computerization and digital informatization is low. Many books are not classified specifically, which makes a lot of scattered materials on many shelves. In the actual investigation of the library, it is found that the number of books with special collections is small and not updated in time. This leads to inefficient use of books. At present, the application of modern Internet information technologies such as books and multimedia in libraries is still relatively low and lacks effective utilization. Especially, the research on network and big data technology in the library is rare. Many scholars believe that the reading amount and quality of books and periodicals are unstable. At the same time, some people pointed out that the current construction of document IR in China needs to be improved [11-12].

The existing problems can be summarized in the following aspects:

The talent team structure is unreasonable. The proportion of librarians outside the four majors of medicine, graphics and information, computer and foreign language is the largest. Librarians are unreasonable in terms of professional knowledge composition and education level. The fund input of the library is insufficient. The annual investment shall not exceed 1 million yuan. The level of network electronization is low. The modernization level of academic libraries (rooms) is not high, and the service means are relatively backward. There are few characteristic databases built. Only a few libraries have special book resources. The level of information retrieval is not high. Readers have a general grasp of retrieval methods, which will restrict users' use of IR in our provincial medical library. The user's information environment is poor. The information environment of the library is ordinary, which can only meet part of the information needs of some users, and the information environment of users is relatively poor. The utilization rate of IR in academic libraries is not high. The utilization rate of library electronic resources is not high. There are problems of repeated construction and waste of IR. The obstacles to users' access to literature mainly come from the poor resource sharing ability of information institutions, followed by the lack of unified management of IR [13-14].

2.3. Countermeasures for Library Information Resources Construction

This paper has the following suggestions for the CL literature IR, and carries out them one by one:

To urge the superior leaders to attach importance to the role of the library. We will speed up education reform, and conduct a multifaceted and comprehensive investigation and assessment of local education and cultural level. The allocation of library staff should also be based on their professionalism and ability. Strengthen the leadership and management of the library and establish a unified management system. The library collection resources should be socialized, and the traditional concepts should be completely abandoned. The ideas and understanding should be unified on the road of co construction, sharing and coordinated development of IR. There should be a leading organization and an information center to organize and lead the sharing of resources. Establish a business training system, and timely learn and absorb advanced library management technologies and methods. Librarians should improve their comprehensive cultural quality, be able to establish a good image among readers, and play a role as teachers. The scientific research achievements of the field visit library are intuitive, easy to understand and effective [15-16]. Librarians should provide quality services and serve readers wholeheartedly. Try to expand the interview space. The library should attach importance to the education and training of interviewers, constantly update knowledge and improve the knowledge structure. Librarians should adjust their acquisition strategies and use limited funds to collect high-quality and diversified library resources. Librarians should adjust the collection structure to meet the needs of the new curriculum reform. Establish a diversified library system. The digitalization of IR will provide rich curriculum resources for the new curriculum, and can be conveniently and quickly used in education and teaching. The IR are networked, relying on the network and providing all-round services. In addition, the construction of electronic reading room is also an important measure of modern library management [17-18].

2.4. Use of Machine Learning in Library Information Resources

With the rapid development of information science, human knowledge has shown an explosive growth. A large number of data are collected and stored in many fields every day. The traditional simple statistical methods have been unable to meet people's needs. Machine learning, as a tool which takes data as the research object and aims to discover rules and associations, is born from this.

In the construction of special collection document IR of the library. The following principles should be followed.

People oriented and humanistic spirit. It is a special and important topic for librarians. Therefore, in order to better serve the readers and enhance their own strength and competitiveness, we must adhere to the concept of humanistic development to carry out internal machine learning and management of libraries. At the same time, we should pay attention to maintaining the working environment of librarians and the hardware facilities and equipment in the library, and put forward suggestions and measures to ensure their normal operation, so as to make the resource construction of the library more scientific and reasonable. Under the condition of making full use of the existing conditions and combining the special collection of literature IR, a database with its own characteristics should be established. Select appropriate topics for data storage according to different retrieval requirements and specific situations. The combination of the principle of information quality first, the principle of gradual progress and the requirement of classification clarity is required. For books and periodicals, we should not blindly invest capital, manpower and material resources to develop new technologies and materials.

3. Investigation on the Construction of Library Special Collection Document Information Resources

3.1. Investigation on Library Literature Resources

Library is the gathering place of IR, whose main functions include searching, searching and using. At present, there are a large number of scattered documents in the Chinese book market. In the context of the Internet plus era, traditional journals and other paper materials have gradually been electronic. Therefore, using machine learning technology to build a model can help us solve these problems. Literature, map and information, computer and foreign language majors have the largest proportion of librarians in other majors. This sign shows that the human resources of the library (room) are unscientific in the professional composition.

3.2. Questionnaire Design

This paper focuses on the construction of special collection literature IR of the library to carry out relevant questionnaires. The content of the questionnaire involves the following aspects: First, the relevant information of the person filling in the questionnaire, age, profession and occupation. Secondly, it investigates the CL special collection literature information, and describes the opinions on the construction of literature IR. Then it expounds the problems of the special collection of information in the current library. Finally, leave the relevant suggestions on the construction of the library to the person filling in the questionnaire.

3.3. Questionnaire Process

The survey lasted 2 weeks in total. The questionnaire was conducted near a local library. A total of 300 questionnaires were collected through online and offline questionnaires. Among them, 98 offline questionnaires and 202 online questionnaires were collected. After sorting out, 270 valid questionnaires were obtained, and the effective questionnaire rate reached 90%. Finally, the statistics method is used to sort out the data.

4. Analysis of Survey Results

4.1. Problems of Library's Special Collection of Literature Information

According to the sorting out of the questionnaire survey results, the problems of library special collection documents mainly include the following aspects: low degree of digitalization, low level of information retrieval, few special collection databases, and low utilization rate of resources. The investigation of specific problems is shown in Table 1:

Table 1. The problem of the library's special collection of literature and information

	Agree	Disagree	Neutrality
Low degree of digitalization	65%	13%	22%
Low information retrieval level	60%	10%	30%
Few special databases	65%	15%	20%
Low resource utilization	70%	10%	20%

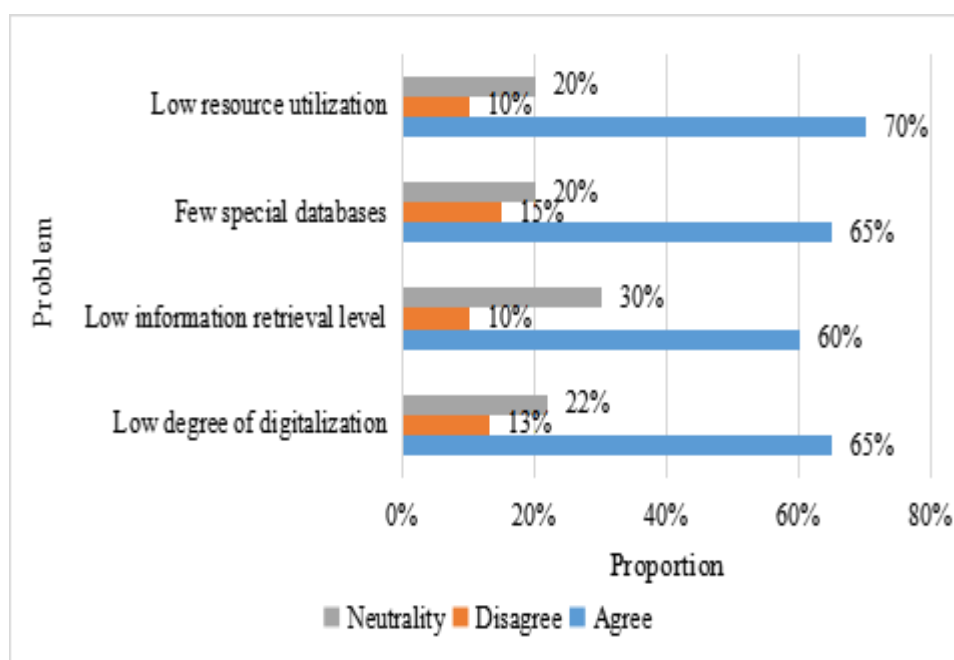


Figure 2. The problem of the library's special collection of literature and information

As shown in Figure 2, we can see that 70% of the people think that the low utilization rate of the resources of the library's special collection of literature information is the primary problem. More than 60% of people think that the information retrieval level of library's special collection literature information is not high, so the utilization rate of library's IR is low.

4.2. Suggestions on the Construction of Library Special Collection Literature Information

In view of the problems existing in the construction of special collection document information in the library, the author of the questionnaire put forward suggestions, including the professional quality of librarians, research on special collection documents, resource sharing technology, financial support, etc. See Figure 2 for details:

Table 2. Opinions on the construction of special library collection of literature and information

	Agree	Disagree	Neutrality
Professional quality of library attendants	30%	20%	50%
Research on special literature	60%	10%	30%
Resource sharing technology	65%	13%	22%
Funding support	68%	6%	26%

As shown in Figure 3, we can find that among the investigators, they believe that financial support can improve the CL special collection literature information. Only a few people think that the professional quality of librarians is harmless. They think that the special collection literature is to study the special collection resources, so that there may be more special collection databases.

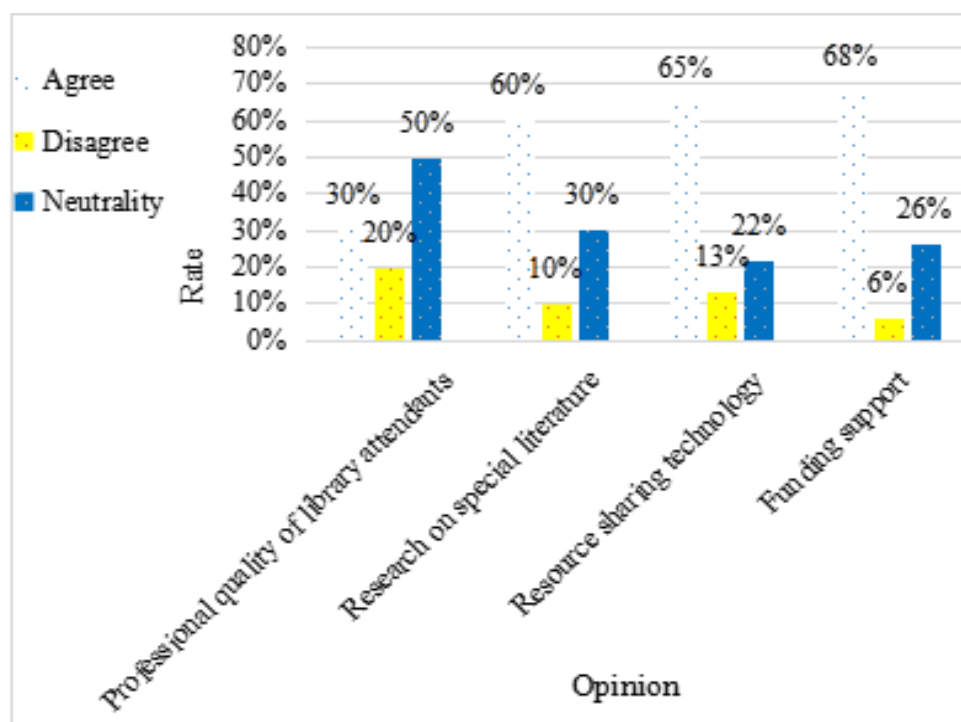


Figure 3. Opinions on the construction of special library collection of literature and information

5. Conclusion

Through the study of library IR, this paper analyzes the existing problems of library special collection literature information, and puts forward corresponding countermeasures in combination with the actual situation. Through the writing of this paper, we learned a series of drawbacks such as insufficient, incomplete and empty literature resources in the library. And according to the current situation of book resource construction, the relevant results are obtained. There are two patterns that can be applied in this field. One is to use internet technology to build a web-based database to achieve retrieval and classification functions. The other part is to take literature information as the research object, and use pictures or videos as the means to build web theme pages to enhance the attractiveness of literature. However, the survey found that people are not skilled and extensive in the use of electronic libraries, so we should strengthen people's search ability and knowledge seeking psychology.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

- [1] Mar ú Luisa Alvite D éz, Leticia Barrionuevo: *Confluence between Library and Information Science and Digital Humanities in Spain. Methodologies, Standards and Collections. J. Documentation* 77(1): 41-68 (2021). <https://doi.org/10.1108/JD-02-2020-0030>
- [2] Suzanna Hall: *Fashion Torn Up: Exploring the Potential of Zines and Alternative Fashion Press Publications in Academic Library Collections. Libr. Trends* 70(1): 51-71 (2021). <https://doi.org/10.1353/lib.2021.0013>
- [3] Melissa Chiavaroli: *Ditching Dewey: Take Your Collections from Enraging to Engaging and Position Your Library for 21st Century Success. Public Libr. Q.* 38(2): 124-146 (2019). <https://doi.org/10.1080/01616846.2018.1555786>
- [4] Julie Linden, Sarah Tudesco, Daniel Dollar: *Collections as a Service: A Research Library's Perspective. Coll. Res. Libr.* 79(1): 86-99 (2018). <https://doi.org/10.5860/crl.79.1.86>
- [5] Melissa Terras, James Baker, James Hetherington, David Beavan, Martin Zaltz Austwick, Anne Welsh, Helen O'Neill, Will Finley, Oliver Duke-Williams, Adam Farquhar: *Enabling Complex Analysis of Large-Scale Digital Collections: Humanities Research, High-Performance Computing, and Transforming Access to British Library Digital Collections. Digit. Scholarsh. Humanit.* 33(2): 456-466 (2018). <https://doi.org/10.1093/llc/fqx020>
- [6] Makiba J. Foster: *Navigating Library Collections, Black Culture, and Current Events. Libr. Trends* 67(1): 8-22 (2018). <https://doi.org/10.1353/lib.2018.0022>
- [7] Umme Habiba, Md. Emdadul Islam: *Assessing the Quality and Credibility of Scholarly Information Resources: An Investigation into Information Behaviours of University Faculty Members. Electron. Libr.* 40(3): 237-255 (2020). <https://doi.org/10.1108/EL-03-2021-0069>
- [8] Nancy M. Lorenzi, William W. Stead: *NLM and the IAIMS initiative: Cross-Institutional Academic/Advanced Systems Contributing to the Evolution of Networked Information and Resources. Inf. Serv. Use* 42(1): 29-38 (2020). <https://doi.org/10.3233/ISU-210134>
- [9] Stacy Tessler Lindau, Jennifer A. Makelarski, Emily Abramsohn, David G. Beiser, Kelly Boyd, Elbert S. Huang, Kelsey Paradise, Elizabeth L. Tung: *Sharing Information about Health-Related Resources: Observations from a Community Resource Referral Intervention Trial in a Predominantly African American/Black community. J. Assoc. Inf. Sci. Technol.* 73(3): 438-448 (2020). <https://doi.org/10.1002/asi.24560>
- [10] Christina Kamposiori, Claire Warwick, Simon Mahony: *Embedding Creativity into Digital Resources: Improving Information Discovery for Art History. Digit. Scholarsh. Humanit.* 37(2): 469-482 (2020). <https://doi.org/10.1093/llc/fqab088>
- [11] David W. Lewis, Tina Baich, Kristi L. Palmer, Willie M. Miller: *The Efficient Provision of Information Resources in Academic Libraries: Theory and Practice. Libr. Trends* 70(3): 323-354 (2020).
- [12] Kamil Korzekwa, Zbigniew Puchala, Marco Tomamichel, Karol Zyczkowski: *Encoding Classical Information Into Quantum Resources. IEEE Trans. Inf. Theory* 68(7): 4518-4530 (2020).
- [13] Helfried Brunner: *CIREN 2021: Bericht über die Session 4 - Distributed Energy Resources and Efficient Utilization of Electricity. Elektrotech. Informationstechnik* 138(8): 605-606 (2021). <https://doi.org/10.1007/s00502-021-00969-z>
- [14] Bryan Wilder, Sze-Chuan Suen, Milind Tambe: *Allocating Outreach Resources for Disease Control in a Dynamic Population with Information Spread. IJSE Trans.* 53(6): 629-642 (2021). <https://doi.org/10.1080/24725854.2020.1798037>
- [15] Galih Abdul Fatah Maulani, Nizar Alam Hamdani, Sukma Nugraha, Asri Solihat, Teten Mohamad Sapril Mubarak: *Information Technology Resources and Innovation Performance in*

- Higher Education. Int. J. Interact. Mob. Technol.* 15(4): 117-125 (2021). <https://doi.org/10.3991/ijim.v15i04.20193>
- [16] Martin Bichler, Alexander Hammerl, Thayer Morrill, Stefan Waldherr: *How to Assign Scarce Resources Without Money: Designing Information Systems that are Efficient, Truthful, and (Pretty) Fair.* *Inf. Syst. Res.* 32(2): 335-355 (2021). <https://doi.org/10.1287/isre.2020.0959>
- [17] Ashkan Davarpanah, Norshidah Mohamed: *Human Resources Information Systems Implementation and Influences in Higher Education: Evidence From Malaysia.* *Int. J. Asian Bus. Inf. Manag.* 11(3): 65-84 (2020). <https://doi.org/10.4018/IJABIM.2020070105>
- [18] Sarah Hare, Julie Marie Frye, Beth Lewis Samuelson: *Open Pedagogy as an Approach to Introducing Doctoral Students to Open Educational Resources and Information Literacy Concepts.* *Libr. Trends* 69(2): 435-468 (2020). <https://doi.org/10.1353/lib.2020.0041>