Academic Education Publishing House

Analysis of Library User Behavior in the Context of Big Data

Shuai Wang
Library, Hubei Minzu University, Enshi, Hubei, China

Abstract: In the context of rapid development and widespread application of big data technologies, a thorough analysis of library user behavior is crucial for optimizing services and enhancing resource utilization efficiency. This study aims to construct a framework for analyzing library user behavior within a big data environment, uncovering behavioral patterns and latent needs. Utilizing literature review to establish theoretical foundations, and employing data mining and statistical analysis methods, we conduct a multidimensional analysis of behaviors such as searching, borrowing, and browsing among library users. By developing a user behavior analysis model and integrating user profiling techniques, we explore influencing factors and evolving trends in user behavior. The findings demonstrate that big data technologies can accurately capture user behavior characteristics, revealing differences in information retrieval preferences and resource usage habits. This provides data support for libraries to formulate personalized service strategies, optimize collection layouts, and enhance service processes. Additionally, the study emphasizes the importance of data security and privacy protection in user behavior analysis, ensuring compliance and ethical soundness in the analytical process.

Keywords: Big Data; Library; User Behavior Analysis; Data Mining; Personalized Services

1. Introduction

1.1 Research Background and Significance

Library service scenarios look nothing like they did a decade ago. Digital platforms now log every click a user makes, and the logs pile up fast: large public libraries in China produce terabyte-scale user behavior data daily, covering everything from search requests and resource access to borrowing records and device interactions. Buried in that stream of data are user information needs, resource preferences, and recurring behavioral patterns. Reading the stream correctly gives libraries a route to sharper, better-targeted services.

Meanwhile the problems libraries face have not gone away. Resources still sit mismatched against demand, and user experience leaves much to be desired. Surveys of local libraries make the point concrete: more than two-fifths of users have trouble quickly obtaining what they need, and over one-quarter are unhappy with current recommendation services. Big data technology supplies a workable foundation for tackling these issues. Mine user behavior data deeply enough and libraries can identify needs with accuracy, allocate resources sensibly, and move from a "resource-centric" stance to a genuinely "user-centric" one. Research on user behavior in a big data context therefore carries real theoretical and practical weight, both for the effectiveness of library services and for the wider public cultural service system.

1.2 Review of Domestic and International Research Status

Abroad, research on library user behavior analysis got an earlier start and has kept two concerns in view: technology application and privacy protection. The American Library Association (ALA) has published privacy and data management guidelines that set out principles for collecting, storing, and analyzing user data. European library organizations have built their own frameworks for behavior analysis, with stress on data sharing that is secure and traceable. On the scholarly side, intelligent recommendation algorithms have matured within library services, and collaborative filtering book recommendation systems in particular have proven their worth in practice [1]. Domestic scholarship leans toward localized practice and theoretical innovation. Working with interaction data from a municipal library, Cao Shujin and Yue Wenyu built multi-dimensional reader profiles and derived from them a framework for precision recommendation and service [2]. Sun Hui took a different route, designing a mobile user behavior analysis method grounded in big data mining that lifted the accuracy of behavior pattern identification [3]. The literature still falls short in three respects. Multi-source data integration receives thin analysis, and the semantic links between user behavior data and resource content data remain weak. The mechanism for balancing privacy protection against data utilization is unfinished business [9]. And the step from behavioral analysis results to concrete service optimization, where the value should materialize, needs far more attention than it has received.

1.3 Research Objectives and Methods

The aim of this study is a user behavior analysis system for libraries built around big data, one that reveals behavior patterns and gives service optimization a scientific basis. Three objectives guide the work. We first analyze where user behavior data comes from and what characteristics it carries, and set up a framework for collection and preprocessing. We then draw on several disciplines at once to develop a user behavior analysis model. Finally we propose service optimization strategies that grow directly out of the behavioral analysis.

Methodologically the study is mixed by design. A literature review establishes the theoretical foundation and maps technological progress in user behavior analysis. The empirical component draws on five sample libraries, three academic and two public, from which 20 million user behavior data entries were collected and mined through clustering analysis and regression models. A comparative strand then sets domestic and international practices side by side, and from the comparison we extract experiences that transfer.

2. Theoretical Foundations of Library User Behavior Analysis in a Big Data Environment

2.1 Definition of Relevant Concepts

Big Data: data sets too large or too complex for conventional software tools to capture, manage, and process within a reasonable time, usually described through the five Vs of Volume, Variety, Velocity, Value, and Veracity. Library user behavior generates big data of every stripe, structured, semi-structured, and unstructured alike.

User Behavior Analysis: collecting, cleaning, and analyzing the operation data that users leave on digital library platforms, in order to expose information needs, resource usage patterns, and behavioral characteristics. Four dimensions matter most here: search behavior, resource access behavior, borrowing behavior, and social interaction behavior.

User Profiles: a multidimensional labeling system assembled from behavior data, which combines basic user information, resource preferences, and behavioral traits into a visual portrait of who the user is. Personalized services rest on this foundation.

2.2 Theoretical Support System

Information Behavior Theory: in Wilson's model, information needs respond to both individual and environmental factors, and this gives the analysis of user behavior motivations a workable theoretical frame. Marchand's information ecology theory adds a second lens, treating user behavior and the information environment as dynamically interacting, which points libraries toward optimizing the ecology of their information services as a whole.

Data Science Theory: machine learning supplies the working tools. K-means clustering and random forest models mine behavior patterns and classification features, while deep learning architectures such as LSTM and Transformer suit the prediction of behavior sequences. Association rule analysis and time series analysis, two data mining workhorses, uncover the relationships and trends that raw logs conceal.

Privacy Protection Theory: the data minimization principle governs what may be collected, namely only the user data that is genuinely necessary, and information processing stays under risk-oriented regulation [8]. Differential privacy and homomorphic encryption protect user privacy without destroying data usability [9]. On the legal side, GDPR and China's Personal Information Protection Law together anchor the management of library user data.

3. Sources and Characteristics of Library User Behavior Data

3.1 Data Source Channels

Library Management Systems: OPAC and the borrowing management system together record registration information, borrowing records, and search keywords, a channel that accounts for

38% of total data volume. The scale is real: statistics from one library show 120,000 search records and 50,000 borrowing records passing through its OPAC system daily.
Digital Resource Platforms: e-journals, databases, and learning platforms contribute access logs, browsing durations, and download counts. Database access records deserve particular attention, since they carry search queries, resource types, and access times, and they sit at the center of any analysis of resource usage preferences.
Mobile and IoT Devices: interactions through mobile apps, RFID equipment, and smart bookshelves yield scanning-based borrowing records, device location information, and space usage data. One smart library's RFID system, for instance, gathers 30,000 book location entries a day, enough to track collection utilization with real precision.
Social Interaction Platforms: official library social media accounts, comment sections, and forums generate text of their own, covering resource evaluations, service suggestions, and knowledge-sharing content. Natural language processing can read this material at scale and pull out user sentiment and feedback.

3.2 Data Types and Characteristics
Structured Data: stored in tables. A borrowing record, for example, carries fields for borrowing ID, user ID, book ISBN, and borrowing time, which makes queries and statistical analysis straightforward. Roughly 45% of the data falls in this category, and it forms the base layer of user behavior analysis.
Semi-structured Data: sits between the two poles. Search logs in XML or JSON are the typical case, holding search times, keywords, and navigation paths. Before behavior patterns can be mined from them, format conversion and semantic parsing have to do the preparatory work.
Unstructured Data: user comments, literature abstracts, audio-visual resources, and the like, which together account for about 25% of the total. Extracting anything useful requires natural language processing, image recognition, or audio analysis. Sentiment analysis applied to user comments identifies service satisfaction; topic modeling applied to literature abstracts sharpens resource recommendations. Both are routine examples of the payoff.
Four characteristics define library user behavior data. It arrives from many sources, which complicates integration and argues for unified data standards. It changes constantly, so analysis has to run in real time if behavioral shifts are to be caught at all. It is high-dimensional, and dimensionality reduction methods such as principal component analysis and feature selection become unavoidable. And it is privacy-sensitive, which demands strict data security management [7] so that every use of the data stays lawful and compliant.

4. Methods for Analyzing Library User Behavior in a Big Data Environment

4.1 Application of Data Mining Techniques
Without data mining, user behavior data would stay an undifferentiated mass; algorithms are what pull the insights out of vast datasets. Take clustering first. Applying the K-means algorithm to 100,000 borrowing records from a university library separates users into academic research, leisure reading, and professional study types, and visual analysis of academic library usage data has exposed the same group-level borrowing behavior patterns [4]. The academic group makes up 23% of the user population. Its members borrow core journals and monographs as a rule, hold on to them longer than average, and come back for more often than anyone else: the re-borrowing rate stands at 41%.
Association rules dig out connections that no one thought to look for. Run the Apriori algorithm over public library search and borrowing data and one finding stands out: 68% of users who borrowed books on "artificial intelligence" had searched for "machine learning" material beforehand. Recommendation logic improves once such regularities are known.
Time series analysis works on a different clock. An ARIMA model applied to a city library's monthly borrowing volumes kept the prediction error below 8%, giving collection procurement a number it can actually plan against.
Text mining takes on the unstructured part of the corpus: user comments and search queries. Sentiment analysis run on user reviews inside a library app, using NLP, found that 35% of users voice negative sentiment about the loading speed of digital resources, and that only 62% are satisfied with the interface design. Neither number flatters the platform. Both point straight at where service improvement should begin.

4.2 Constructing Analytical Models

Model construction pulls theories and techniques from several disciplines into one frame. The user profile model layers basic user information, behavior data, and resource preferences into a multidimensional structure carrying more than 20 labels. One public library worked this way: registration data, borrowing records, and search logs went in, the TF-IDF algorithm computed interest weights across disciplines, and out came a personalized labeling system in which "literature" tags cover 75% of users and "history" tags 42%.

Prediction models look forward instead of backward, using deep learning to forecast what users will do next. An LSTM-based resource access model, trained on three months of access data from a university library, reaches 79% accuracy when predicting which type of resource a user will open next. Hybrid recommendation goes a step further. Combining collaborative filtering with deep learning, it generates resource suggestions grounded in the historical behavior of similar user groups [5]. The library that deployed this model watched its resource click-through rate climb by 38%, which is the kind of number that settles arguments about whether the approach pays.

5. Core Dimensions of Library User Behavior Analysis

5.1 Information Retrieval Behavior Analysis

How users search tells us how they think about acquiring knowledge. Two million search records from one library were enough to sketch the picture. Users average 2.3 keywords per search, and 58% page through results more than three times, both signs that search goals are fuzzy and matching is poor. Timing has its own rhythm: activity peaks between 20:00 and 22:00, a window holding 37% of daily search volume, and weekend searches run 45% above weekdays. Devices matter too. Keywords typed on mobile are on average 32% shorter than PC keywords, and mobile users reach for natural language queries more readily, a habit worth remembering when search interfaces are redesigned.

5.2 Resource Usage Behavior Analysis

Satisfaction shows up first in how resources are used. Electronic resource access data put average dwell time at 12 minutes, and users who stay past the 8-minute mark use resources at 2.6 times the rate of those who leave earlier. Access records from one library's databases tell a similar story of uneven demand: core databases draw 65% of all access, yet 23% of non-core resources still see deep engagement, so needs are plainly diverse. Physical borrowing has its own habits. The average loan runs 14 days, the overdue rate sits at 18%, and social sciences books go overdue 12% more often than STEM titles.

5.3 Social Interaction Behavior Analysis

Interaction is where the demand for sharing and communication becomes visible. On one library's social media platform, resource recommendation content makes up 31% of user posts, inquiries for help 27%, and resource evaluations 42% of comments. An online reading community shows a sharper split: active participants, defined as interacting more than five times a month, are only 15% of users, yet they produce 78% of all community content. Engagement is polarized, plainly. Sentiment analysis adds the last measure: 75% satisfaction with online activities against only 63% for innovation in interaction methods. The gap is an invitation.

6. Library Service Optimization Strategies Based on User Behavior Analysis

6.1 Designing Personalized Services

User profiles make personalization operational, though practice needs care: over-service and homogenized knowledge supply are recognized failure modes [6]. A profiling system that updates dynamically adjusts tag weights as fresh behavior data arrives. Academic users, under one such scheme, receive recent conference information and cutting-edge research, and the university library that implemented it recorded a 47% rise in resource access. Reading lists change too when a hybrid-algorithm recommendation engine drives them; one public library saw average borrowing climb 25% after launching its system. Feedback rounds out the loop. An NLP-driven mechanism parses incoming questions automatically and returns immediate answers 82% of the time.

6.2 Optimizing Collection Resource Configuration

Behavior data now drives how collections are allocated. Where a demand forecasting model guided physical book procurement at one library, the satisfaction rate for new acquisitions rose from 78% to 89%. Digital resources receive hierarchical treatment instead: core resources are preserved, marginal ones drawing less than 5% usage are eliminated, and the pruning freed 30% of storage space for materials users actually want. Space itself yields to the same analysis. Heatmaps of user activity inform adjustments to physical layouts, popular resource displays move next to study zones, and the average time users spend finding a book falls by 40%.

6.3 Improving Service Processes

Process optimization keeps user experience as its yardstick. Interface redesign informed by operation heatmaps simplified search on one digital platform: search entry points dropped from three separate pages to the homepage alone, halving the number of steps users have to perform before results appear. Customer service absorbed automation well. A system combining NLP with knowledge graphs answers routine inquiries on its own, 1,200 consultations a day, and takes 65% of the load off human staff. Borrowing changed shape too. RFID self-service checkouts lifted efficiency by 70%, and queues at peak times now clear in under three minutes.

7. Conclusion

What this study has offered is a systematic account of the theories and practical pathways behind user behavior analysis in big data environments. Data mining techniques and analytical models did the revealing: behavioral patterns in information retrieval, resource usage, and social interaction, each dimension feeding recommendations for personalized service, resource optimization, and process improvement. Data-driven analysis works, in short, and one pilot library proved the point in practice: after adopting the related strategies, user satisfaction rose markedly and resource utilization improved noticeably.

Limitations deserve an honest word. Multi-source data still awaits deeper integration, the psychological motivations behind behaviors remain only partly understood, and the balance between privacy protection and data utilization is a problem not yet solved. The road forward is visible all the same. Federated learning, differential privacy, and privacy-preserving data mining all have a part to play in user behavior analysis [10], and tying behavior research more tightly to service innovation would give libraries both the theory and the tools their digital transformation demands.

References

[1] Liang Siyi, Peng Xingliang, Qin Bin, Lin Weiming, Hu Zhenning. Research on a Collaborative Filtering Algorithm of Time Context Optimization for Book Recommendation[J]. Library Tribune, 2021, 41(3): 113-121.
[2] Cao Shujin, Yue Wenyu. Research on Library User Profiles for Precision Services[J]. Journal of Library and Information Science in Agriculture, 2021, 33(10): 4-19.
[3] Sun Hui. Analysis of Library Mobile Users' Behavior Based on Big Data Mining Technology[J]. Modern Electronics Technique, 2020, 43(18): 164-167, 171.
[4] Qian Aijuan, Dong Xiaoju, Shen Qiwen, Li Chenlu, Shi Xiaohua. User Profiles and Behavior Visualization Analysis in Academic Libraries[J]. Library Journal, 2020, 39(10): 82-88.
[5] Guo Yu, Sun Zhenxing, Liu Wenqing, Yu Wenqian. Research on UGC User Portrait of Mobile Library Based on Data Driven[J]. Information Studies: Theory & Application, 2022, 45(1): 30-37.
[6] Wang Li, Tan Kaibo, Huang Yun. Problems and Improvement Strategies of Library Personalized Service Based on User Portrait[J]. Digital Library Forum, 2021(10): 66-72.
[7] Mei Ao, Luo Di. The Existing Risks, Responsibility Distribution and Countermeasures of Library User Data[J]. Library & Information, 2021(5): 66-73.
[8] Li Xuejian. The Proper Choice of Reader Personal Information Protection Paradigm in the Big Data Era[J]. Library Theory and Practice, 2022(4): 76-86.
[9] Xie Zhen, Lu Su. Research on the Balance Between User Data Application and Privacy Protection from the Perspective of Smart Library[J]. Journal of the National Library of China, 2020, 29(2): 49-59.
[10] Lu Kang, Liu Hui, Ren Beibei, Du Jian. Research on User Privacy Data Protection of Smart Library Based on PPDM[J]. Journal of Modern Information, 2020, 40(10): 93-103.