Environment. Technology. Resources. Rezekne, Latvia
Proceedings of the 16th International Scientific and Practical Conference. 2025. Volume II, 37-41

Using Bayesian Methods to Predict Users' Information Needs in a Digital Library

Fayzi Bekkamov
Library information systems department Tashkent University of Information Technologies named after Muhammad al-Khwarizmi
Tashkent, Uzbekistan
bfayzi93@gmail.com

Mumin Babajanov
Modern Information and Communication Technologies department International Islamic Academy of Uzbekistan
Tashkent, Uzbekistan
mum1975@mail.ru

Mansur Berdimurodov
Modern Information and Communication Technologies department International Islamic Academy of Uzbekistan
Tashkent, Uzbekistan
mansur_alisherovich@mail.ru

***Abstract*— In digital library, identifying users' information needs and providing them with relevant resources is of critical importance. Users' characteristics, personal preferences, and demographic information play a key role in determining their information needs. Machine learning based methods are widely utilized to recommend resources that align with users' information requirements, significantly enhancing the efficiency of information retrieval systems. The Bayesian method is one of the most effective methods for predicting users' information needs. This approach is based on the principle of conditional probabilities and enables the prediction of future information needs by analyzing users' past search behaviors and interaction patterns. The Bayesian model offers several advantages, including the ability to process large datasets, high classification accuracy, and computational efficiency. Due to these benefits, it is extensively applied in digital library to optimize information retrieval and recommend relevant resources to users. This article examines the fundamental principles of the Bayesian method in the context of digital library, highlighting its advantages and limitations. Furthermore, the effectiveness of this model in identifying users' information needs and recommending appropriate resources is analyzed. The findings of this research contribute to the improvement of digital library, enhancing user search experiences and optimizing resource recommendations.**

Keywords*— *digital library, Bayes methods, personal preferences, information requirements, predict, resource, recommender systems, rating, metadata similarity, probability.

I. Introduction

Information needs are influenced by a combination of users' search histories, demographic attributes, and individual preferences. To effectively predict and address these needs, various machine learning techniques have been employed, among which the Bayes method is widely recognized for its efficacy in probabilistic classification and recommendation systems [1, 2].

The Bayes method, derived from Bayesian probability theory, is a powerful statistical classification technique that is utilized for computing the probability of a user seeking particular information, based on the past data or experiences. The technique is particularly adept at predicting users' information needs because it is developed upon analyzing various variables such as their earlier search histories, patterns, interests, and contextual variables.

The application of Bayesian techniques for information need prediction generally includes a number of key steps:

Data collection and preprocessing: the first task is to gather users' search queries, browsing history, and other available interaction data. Raw data is cleaned and preprocessed to eliminate noise and irrelevant information so that the model learns from meaningful and quality input.

Pattern extraction: in this stage, the system scans the processed data for recurring patterns, behavior, or keywords that indicate the interests of the user and possible information needs. These patterns are relevant as they allow the system to make educated predictions regarding what the user might be interested in, given their past behavior.

Bayesian Classification: the second step is to apply Bayesian classification principles for the categorization of the search queries and interactions. Bayesian classification utilizes a probabilistic method that updates the probability of a user's information needs being met based on new data or evidence. The system calculates the probability that a specific query falls under a specific category or topic, based on previous probabilities and the evidence gathered from previous searches.

Online ISSN 2256-070X
https://doi.org/10.17770/etr2025vol2.8586
© 2025 The Author(s). Published by RTU PRESS
This is an open-access article under the Creative Commons Attribution 4.0 International License.

The use of Bayes methods in information retrieval simultaneously optimizes the relevance and effectiveness of the results. With expectations of what is most likely to be searched or what the user will require, the system is able to make useful suggestions or rule out search results. This, besides enhancing the use experience, helps in the development of intelligent, context-aware search engines and recommender systems that are capable of continuously changing based on users' feedback and behaviour [3, 4].

By leveraging the Bayes method, modern information retrieval systems can enhance user experience by providing tailored content recommendations, improving search result accuracy, and predicting future information-seeking behaviors. This approach has been extensively applied in various domains, including digital libraries, e-commerce, and personalized content delivery platforms, underscoring its significance in the field of information science and artificial intelligence [5, 7].

II. Material and Methods

Its application in library information systems significantly enhances the quality and efficiency of service provision through systematic analysis of user behavior and classification of users according to informational needs. The probabilistic character of this approach is particularly effective in large-scale environments where a great amount of user interaction data can be utilized to determine patterns and preferences more precisely [6].

Central to the Bayes approach is Bayes' Theorem, a foundation idea in probability theory that allows for the computation of the probability of an event given prior knowledge and new evidence. In information systems, this would involve estimating the probability that an observed user query or activity reflects a particular information category or need.

By representing uncertainty and revising beliefs in a dynamic manner as new evidence arrives, the Bayesian model allows for adaptive decision-making and sophisticated user profiling. This allows library systems to dynamically adjust resource recommendations, improve search result relevance, and tailor the user experience as a whole [6, 8, 9]

The general form of the Bayes method is:

$$P(C_i \mid X) = \frac{P(X \mid C_i) \cdot P(C_i)}{P(X)} \qquad (1)$$

where,

$P(C_i|X)$ – is the probability that the user belongs to category C_i given the features X;

$P(X|C_i)$ – is the likelihood of observing features X if the user belongs to category C_i;

$P(C_i)$ - is the prior probability of users belonging to category C_i;

$P(X)$ - is the overall probability of users exhibiting features X.

To predict a user's information needs, the probability calculations are based on the following user characteristics:

X_1 - genres of books read by the user;

X_2 - keywords searched by the user;

X_3 - authors searched by the user;

X_4 - topics read by the user.

If C represents user need categories (e.g., dissertations, scientific articles, theses), the probability is calculated using:

$$P(C_t \mid X_1, X_2, X_3, X_4) = \frac{P(X_1, X_2, X_3, X_4 \mid C_t) P(C_t)}{P(X_1, X_2, X_3, X_4)} \qquad (2)$$

Assuming that features are independent of each other, the formula simplifies to:

$$P(C_1 \mid X_1, X_2, X_3, X_4) = \frac{P(X_1 \mid C_1) P(X_2 \mid C_1) P(X_3 \mid C_1) P(X_4 \mid C_1) P(C_1)}{P(X_1) P(X_2) P(X_3) P(X_4)}$$

(3)

Here, each feature is computed independently, simplifying the method and ensuring efficiency.

If the feature set is $X = \{x_1, x_2, \dots, x_n\}$ then:

$$P(X|C_k) = P(x_1|C_k) \cdot P(x_2|C_k) \cdot \dots \cdot P(x_n|C_k) \qquad (4)$$

The Bayes method calculates probabilities for each class C_k and selects the class with the highest probability as the predicted classification:

$$C_{predict} = \arg max = P(C_k|X) \qquad (5)$$

In an library information system context, a scientific researcher searching for academic papers or textbooks can be classified using the Bayes method as follows: Let

C_1 - scientific articles;

C_2 - textbook;

The user's query consists of:

x_1- scientific, x_2- article, x_3- research, x_4- news.

The classification of search results using the Bayes method is presented in Table 1.

TABLE 1 CLASSIFICATION USING THE BAYES METHOD

Keyword (user's query)	C_1 (class)	C_2 (class)
x_1	40	10
x_2	35	25
x_3	50	5
x_4	10	40
Total	135	80

The class probabilities are calculated as:

$$P(C_1) = \frac{135}{135 + 80} \approx 0{,}63$$
$$P(C_2) = \frac{80}{135 + 80} \approx 0{,}37$$

The likelihoods for each keyword given the class are computed as follows:

$$P(x_1|C_1) \approx 0{,}296$$
$$P(x_1|C_2) \approx 0{,}125$$
$$P(x_2|C_1) \approx 0{,}259$$
$$P(x_2|C_2) \approx 0{,}3125$$
$$P(x_3|C_1) \approx 0{,}370$$
$$P(x_3|C_2) \approx 0{,}0625$$
$$P(x_4|C_1) \approx 0{,}074$$
$$P(x_4|C_2) \approx 0{,}5$$

Using the Bayes method, we compute:

$$P(C_1|X) = P(C_1) \cdot P(x_1|C_1) \cdot P(x_2|C_1) \cdot P(x_3|C_1) \cdot P(x_4|C_1) \approx 0{,}0019$$

$$P(C_2|X) = P(C_2) \cdot P(x_1|C_2) \cdot P(x_2|C_2) \cdot P(x_3|C_2) \cdot P(x_4|C_2) \approx 0{,}00045$$

The query is classified under “scientific article”.

The Bayesian method is a probabilistic framework for user information need determination that follows statistical inference to analyze and categorize their queries. Such an approach fundamentally centers on computing the posterior probability that a given query corresponds to a certain category of information based on prior knowledge and observed data. Therefore, by carefully estimating the likelihood of other user intents, the Bayesian model can classify queries with reasonably high accuracy.

III. Results and discussion

This probabilistic inference process, nevertheless, enables the system to incorporate historical user activity and contextual attributes within its decision-making. It is therefore not merely looking for keywords associated with a query but contextualizing queries within a general informational framework, enabling more subtle and useful categorization. This then offers a way of fetching highly relevant resources and can also be utilized in the development of smart, user-oriented information retrieval systems that can adapt to various and ever-changing needs of users.

Figure 1 illustrates a comprehensive information model specially made for predicting user information needs in the context of a library information system based on the ideas of Bayesian inference. The model operates via a systematic sequence of processes that begin with the systematic inspection of user-generated search queries. It proceeds to screen and retrieve the most pertinent information resources from a vast and heterogeneous digital library.

Through constant observation and interpretation of user activity, the model applies probabilistic inference to predict the likelihood of specific information needs. Through the inclusion of Bayesian inference techniques, the system can iteratively improve its predictions based on new data so that it can dynamically adapt to the changing user preferences, behaviours, and context situations of individual users [10].

Therefore, the system not only enhances the efficiency and relevance of information retrieval but also significantly increases the accuracy and individualization of resource recommendations. This dynamic approach ensures that users are presented with on-time and contextually appropriate information, hence improving the overall effectiveness of the digital library environment for meeting user needs [11].

One of the main strengths of the Bayesian classification method over other statistical and machine learning classification methods lies in its extensive computational efficiency. The Bayesian process is particularly good for large-sized information environments due to its capacity to process and analyze enormous text databases with fantastic speed and with minimal computational power.

This is because the algorithm relies on relatively simple probabilistic computation, which allows for fast training and classification even for high-dimensional data. Furthermore, the probabilistic nature of the Bayesian method to inference provides a principled approach to handling uncertainty and missing information. By making posterior probabilities dependent on prior evidence and knowledge, the model provides for sound and confident decision-making, even in challenging and dynamic information retrieval situations [12].

These properties make the Bayesian approach especially suitable for application in massive information retrieval systems, such as digital libraries, web search engines, and recommendation websites, where speed, scalability, and timely response to noisy user activities are critical to system performance and user satisfaction.

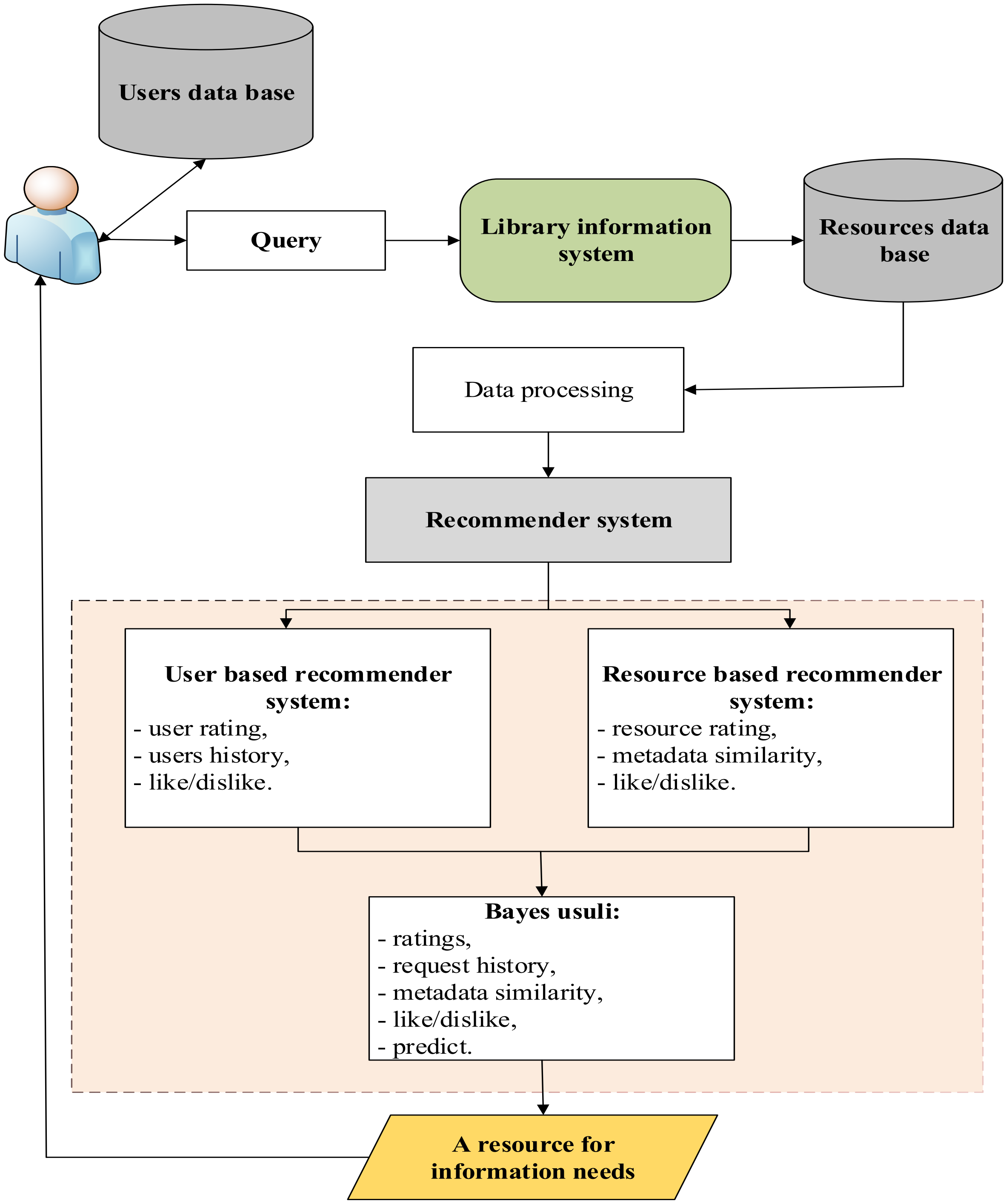

Fig. 1. An information model for identifying and recommending information needs of users in the library system.

Online ISSN 2256-070X
https://doi.org/10.17770/etr2025vol2.8586
© 2025 The Author(s). Published by RTU PRESS
This is an open-access article under the Creative Commons Attribution 4.0 International License.

IV. CONCLUSIONS

The application of Bayesian methods in library information systems significantly enhances the accuracy and efficiency of information retrieval and recommendation processes. By leveraging conditional probability principles, Bayesian models can effectively predict users' information needs based on their past search behaviors and interaction patterns. This method enables the classification of user queries with high precision, improving the personalization of content recommendations and search result relevance.

One of the key strengths of the Bayesian approach is its computational efficiency, allowing it to process large-scale datasets rapidly while maintaining high classification accuracy. The ability to independently compute features simplifies the classification process and ensures scalable implementation in library information systems environments. Furthermore, Bayesian inference dynamically adapts to changing user behaviors, making it a robust tool for optimizing information retrieval systems.

Through probabilistic modeling, the Bayesian method enhances user experiences by providing more accurate and tailored content recommendations. Its effectiveness has been demonstrated in various domains, including library information systems, e-commerce platforms, and personalized content delivery systems. Given its advantages, the Bayesian method remains a critical tool in the advancement of modern information retrieval and recommender systems, contributing to more effective and user-centered library services.

V. REFERENCES

[1] M. Condliff, D. Lewis, D. Madigan, and C. Posse, Bayesian mixed-effects models for recommender systems. ACM SIGIR Workshop on Recommender Systems: Algorithms and Evaluation, 1999

[2] W.Jin, User interest modeling and collaborative filtering algorithms application in English personalized learning resource recommendation. Soft Comput 2023. https://doi.org/10.1007/s00500-023-08700-0

[3] M. Rakhmatullaev, S. Normatov, F. Bekkamov, Fuzzy Relations Based Intelligent Information Retrieval for Digital Library Users, Environment. Technology. Resources. 14th international scientific and practical conference. June 15-16, 2023, Rezekne Academy Of Technologies, Rezekne, Latvia

[4] A.Rustamov, F.Bekkamov, Recommender systems: an overview, Scientific reports of Bukhara State University, 2021/3(85)

[5] C.Yang, Q. Fan, T. Wang et al., RepoLike: a multi-feature-based personalized recommendation approach for open-source repositories. Front Inf Technol Electron Eng 20(02):86–101, Springer 2019

[6] K. Miyahara, and M. J. Pazzani, Collaborative filtering with the simple Bayesian classifier. Pacific Rim International Conference on Artificial Intelligence, 2000

[7] F. Bekkamov, Y. Sharifov, Extended User Profiling Approaches for Recommendation Systems, Environment. Technology. Resources, 15th International Scientific and Practical Conference. June 27-28, 2024, ‘’Vasil Levski’’ National Military University, Veliko Tarnovo, Bulgaria

[8] S. Farshidi, K. Rezaee, S., Mazaheri . et al. Understanding user intent modeling for conversational recommender systems: a systematic literature review. 2024. https://doi.org/10.1007/s11257-024-09398-x

[9] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, BPR: Bayesian personalized ranking from implicit feedback. Uncertainty in Artificial Intelligence (UAI), pp. 452–451, 2009.

[10] Ф.А. Беккамов, М.А. Рахматуллаев, Система поиска научной и образовательной информации с анализом сведений о пользователях, Информатизация образования и методика электронного обучения: цифровые технологии в образовании, Материалы VIII Международной научной конференции, 24-27 сентября 2024 г. Красноярск

[11] M. Rakhmatullaev, S, Normatov, F. Bekkamov, Tavsiya etish tizimlarining umumiy tahlili, Raqamli Transformatsiya va Sun’iy Intellekt ilmiy jurnali, Volume 1, Issue 4, December 2023

[12] T. Ha, S. Lee, Item-network-based collaborative filtering: a personalized recommendation method based on a user’s item network. Inf Process Manage 53(5):1171–1184, Springer 2017