

# Research on Image Fast Retrieval Application based on Bayesian Network in Hybrid Space

Lina Han<sup>1,2</sup>, Nan Shi<sup>2</sup>

<sup>1</sup>Shaanxi Xueqian Normal University, Xi'an, Shaanxi, 710100, China

<sup>2</sup>Xi'an Shiyou University, Xi'an, Shaanxi, 710065, China

**Keywords:** Image Retrieval Application, Bayesian Network, Hybrid Space

**Abstract:** The Bayesian network is based on statistics and is a method of data mining technology. Essentially, the Bayesian network is a directed, non-cyclical chart model that visually expresses the dependencies between multiple variables. It describes the causal relationship between nodes through a directed non-cyclic graph, and describes the degree of close relationship between nodes through a conditional probability distribution table. Moreover, the Bayesian network can effectively combine prior knowledge with existing data, making the network's reasoning results more reasonable. This advantage is more pronounced in Bayesian networks, especially when current data is scarce or difficult to obtain.

## 1. Introduction

Among various information services based on the Internet, the document data repository is one of the most common and frequently used sources of information. As documents continue to grow and become larger, the retrieval of document databases becomes more and more difficult. When a user wants to find a document related to a topic, the search tool uses its powerful search capabilities to recommend a large number of documents to the user without concern whether the documents are really needed by the user. From the user's point of view, they must make their own choices in these recommended documents. This part of the work is done manually by the user, which not only makes them feel exhausted, but also wastes a lot of time and expense. Moreover, when users search for documents and input query words, they are often not accurate enough, which has a certain impact on the results of the search, which directly leads to the efficiency of retrieval. Therefore, users are eager for a new type of search tool that can understand the user's actual intentions, filter redundant information, and accurately expand the search terms.

## 2. Intelligent information retrieval

Information retrieval based on artificial intelligence is a new type of retrieval method which has emerged in recent years. It integrates knowledge and advanced technology in the fields of expert system, natural language understanding, user model, pattern recognition, database management system and information retrieval. For a distributed information space such as the Internet, the artificial intelligence method is a better method to realize personalized retrieval. It can replace the complicated tasks of information collection, filtering, clustering and fusion, and guide users in the Internet. Conduct a more efficient search.

At present, some foreign scientific research departments, colleges and universities, commercial companies are studying intelligent network information retrieval, and have developed some successful products. Such as: Arthur Andersen's FSA and Eloise systems embedding domain-specific knowledge and using proven natural language understanding techniques; IBM's rules and knowledge, using heuristic strategies and a simple natural language Globenet system; developed by the University of Chicago Intelligent search engine FAQFinder based on "question library" with Q&A; WebWatcher based on machine learning-based intelligent system developed by Carnegie Mellon University; and dedicated intelligent software WeDoggie (CMU) based on user

query behavior and interest to find specific information News Weeder, Firely, NewsFinder, etc. Domestic research on intelligent network information retrieval has also been carried out, such as the WebAccess system developed by Nanjing University, which uses techniques such as machine learning, natural language processing, hypertext, etc.; PINS system and Bookmark system developed by Tsinghua University They can automatically collect and record user habits and interests, and track user information needs. The first letter intelligent search engine adopts the "network-to-network" technology; the "web code" AISS system based on Chinese information processing technology such as Chinese grammar and word context semantics; the square intelligent search engine based on user personality requirements. At present, most of the domestic intelligent network information search systems only support simple natural language understanding and concept retrieval, and few researches on machine learning, intelligent agents, information mining and other technologies.

Online intelligent information retrieval is an effective means to help people quickly obtain information. However, the existing systems still have some shortcomings or deficiencies. For example, the non-personalized retrieval method has a poor ability to adapt to changes in user interest, the user interaction with the retrieval system is relatively monotonous, and the ability to adapt to information source information changes is lacking.

### **3. Bayesian network basic concepts**

The ground-breaking work of the Reverend Thomas Bayes (1702-1761) school is Bayesian paper "Comments on Solving Probability Problems." Perhaps it is his own feeling that there is still imperfection in his doctrine. This paper was not published before his death, but was published by his friends after his death. The famous mathematician Laplace P.S. used the Bayesian method to derive important "sequential laws". Bayesian methods and theories were gradually understood and valued. However, because there were still many imperfections in the theoretical and practical applications of the Bayesian method at that time, it was not universally accepted in the 19th century. At the beginning of the twentieth century, Italy's B. de Finetti and Britain's Jeffreys H. all made important contributions to the Bayesian theory. After the Second World War, Wald A. proposed the statistical decision theory, in which the Bayesian school occupied an important position; the development of information theory also made new to the Bayesian school. contribution. In 1958, Biometrika, the oldest statistical journal in the UK, re-published Bayesian papers. In the 1950s, Robbins H. represented the combination of empirical Bayesian methods and classical methods. The extensive attention of the statistical community, this method quickly showed its advantages and became a very active direction. With the development of artificial intelligence, especially machine learning, data mining, etc., it provides a broader space for the development and application of Bayesian theory. The connotation of Bayesian theory has also changed a lot. The Bayesian network in the 1980s was used for the knowledge representation of expert systems. In the 1990s, the Bayesian network that could be learned was further studied for data mining and machine learning. In recent years, Bayesian learning theory has emerged in an endless stream, covering most areas of artificial intelligence, including causal reasoning, uncertainty knowledge representation, pattern recognition and cluster analysis. And there have been organizations and academic journals ISBA that specialize in Bayesian theory.

In uncertainty reasoning, Bayesian network belongs to a model-based connotation method. Compared with rule-based extension method, its advantage is clear semantics, and the disadvantage is computational complexity. In simple terms, it provides a model representation of domain-specific knowledge and several learning and reasoning mechanisms based on this model for building models and answering queries related to knowledge in these areas, and assisting on this basis. Forecasting, decision making, and analysis.

### **4. Bayesian network application**

At present, the research on Bayesian network mainly focuses on three aspects: Bayesian network

inference, Bayesian network learning and Bayesian network application. We study the information retrieval model, mainly using Bayesian network reasoning and application, that is, the process of computing with Bayesian network model. Bayesian network has gradually replaced the early rule-based methods in the field of uncertainty reasoning of expert systems due to its expansion and improvement of reasoning mode and rigorous mathematical foundation. It has been widely used in assisted intelligent decision-making, data fusion, Pattern recognition, medical diagnosis, text comprehension, prediction, information classification, data mining, system control, etc., and the emergence of the organization and academic journal ISBA specializing in Bayesian theory. Bayesian network-based application software can be roughly divided into two types: expert systems for specific fields and general Bayesian network development tools.

Bayesian's outstanding performance in uncertain knowledge representation and reasoning provides a powerful tool for other areas of research in artificial intelligence. In pattern recognition, the classifier constructed using the simplest Bayesian network, Naive Bayes, has the same classification ability as the classic C4.5 classifier. The classifiers modified by Bayesian network, such as tree-enhanced Naive Bayes classifier and general-purpose network classifier, have better classification ability than C4.5 classifier. In data mining, the Bayesian network learning algorithm can be used to analyze and sort data offline, or to update data online. In recent years, with the deepening of research, the use of Bayesian networks to learn the implicit causality in data has made great progress, which provides a more powerful tool for knowledge mining. At the same time, a number of expert systems based on Bayesian networks have also been established. Such as the earliest Path Finder system, which is a medical system for the diagnosis of lymphatic diseases, it can diagnose more than 60 diseases and involves more than 100 kinds of symptoms; the Internist-I system developed later is also a medical diagnosis system, which can diagnose Up to 600 common diseases. Although most of these systems are now only used for teaching and have not been used for clinical diagnosis of medicine, the Bayesian network-based expert system has gradually been accepted. In the field of business applications, a group of companies represented by Microsoft have applied Bayesian networks to their own products. As early as 1993, Bill Gates, the president of Microsoft, known for his keen business acumen, saw the potential business opportunities of Bayesian Networks, and received the Microsoft's outstanding players in the Bayesian network, Eric Horvitz, David Heckerman and Jack Breese. Under the account, jointly develop the application of Bayesian network. In 1995, Microsoft introduced the first Bayesian-based expert system, a website for early childhood care, Microsoft On Parent ([www.onparenting.msn.com](http://www.onparenting.msn.com)), which allows parents to diagnose their own diseases. In 1996, Microsoft provided Online Trouble Shooters, an online diagnostic for printer failure in Win95, based on a Bayesian network with more than 50 nodes, involving four to five major symptoms of the printer, each symptom can be provided The cause of more than a dozen faults. In addition, the Bayesian network-based help system has also been used in Microsoft Office, one of Microsoft's main products. This tool, called Helping Elf, is based on the simplest Bayesian network, Naive Bayes' inference system. It can provide users with different help topics in a timely manner based on past experience. For example, when the user's operation seems to have no clue, the system can conclude that the user needs help. When a user operates on a form, the system will promptly give a prompt about the form of the form.

## 5. Conclusion

Intelligent and personalized search has always been a concern of network users, and it is also one of the focuses of professionals in related fields who have devoted themselves to research and exploration in recent years. The increasing maturity of Bayesian network research has led to further development of intelligent retrieval technology. The information retrieval model proposed in this paper is based on the Bayesian network's reasoning and combines the mining rules of association rules to help users of the document database to perform simple, fast and accurate queries. However, due to the ambiguity and randomness of user behavior, it is impossible to describe and explain user behavior without bias. Therefore, there are still some errors or differences in the model, and the improvement of these errors is still The development and improvement of data mining technology

and other related technologies are waiting. Since I have not studied the Bayesian network for a long time and the level is limited, I would like to ask you for your valuable comments.

### **Acknowledgements**

Fund Project: Shaanxi Provincial Education Science “13th Five-Year Plan” SGH17H195, Shaanxi Natural Science Project 2017JM6086

### **References**

- [1] Xu Jianmin, Tang Wansheng. An Extended Belief Network Retrieval Model Based on Synonym of Query Term [J]. Computer Engineering, 2007, 33(10): 28-30.
- [2] Xu Jianmin, Tang Wansheng, Chen Zhenya. Application of Bayesian Network in Information Retrieval [J]. Journal of Hebei University: Natural Science Edition, 2007, 27(1):93-98.
- [3] ZHU Song, GAO Yanping. Bayesian Network Model Extension for Information Retrieval [J]. Inner Mongolia Science and Technology and Economy, 2009(22): 90-91.
- [4] WU Shufang, XU Jianmin, SUN Xiaolei. Topic recognition model based on Bayesian belief network [J]. Application Research of Computers, 2014, 31(3):792-795.
- [5] WU Shufang, XU Jianmin. Timing Weights in Dynamic Topic Tracking [J]. Computer Science, 2015, 42(2): 233