

Application of Computer Data Mining Technology in Archives Information Management System

Guiqi Liu

Nanyang Institute of Technology, Nanyang, Henan, 473000

Keywords: archive management; information system; computer; data mining technology

Abstract: In the process of organization operation management, archives are an important part of information resources. It is not only a witness and record for the history of organizational development, but also provides a reference for the future development of the organization. At present, many organizations have realistic problems such as backward management, low efficiency, and low level of informatization in file management. They cannot conduct comprehensive and in-depth analysis of information resources. The management content of file management personnel is only archives, statistics, inquiries, etc. With the continuous development of computer technology, the application of data mining technology in information processing is more and more extensive. This paper proposes an archive information management system based on data mining technology, in order to improve the utilization efficiency of information resources.

1. Overview of Data Mining Technology

Data mining refers to extracting implicit information from a large amount of accompanying data, and then applying information resources to the knowledge processing system through technology integration, and mining unknowns from the incomplete, unclear, random data, but there are Value and data that can be used directly to improve the efficiency of the use of information resources. Based on the technical layer to analyze the data mining technology, the data mining technology belongs to the category of commercial data processing technology. It constructs a systematic analysis model and processing mechanism to integrate, analyze and transform commercial data, which can be optimized based on the technical layer to face business decisions. . Data mining technology mainly uses data sources and data preprocessing tools to shape and update data. Data processing is performed through techniques such as correlation analysis algorithm, decision tree analysis module and cluster analysis module to meet information processing integration, time-varying, and non-easy.

2. The Form of Data Mining Technology

Generally, data mining includes two forms, namely, descriptive narrative type and predictive inference type. The former is based on the law describing the existence of data, combined with subtle data, representative and profound meanings of the data resources. The content of the representation is analyzed, and the results of the final data analysis are highly summarized and abstract. The predictive inference type is to obtain the same kind of attributes of the same kind of object after in-depth analysis of the actual data, and then predict the future development trend of the kind of things from another angle. Descriptive narrative and predictive inferred data mining will be combined with induction, related rules, rough sets and other forms. In the whole data mining technology, induction will play a decisive role in the attributes and advantages and disadvantages of data collection. It divides and analyzes the database attribute group elements. The training set formed by data is targeted in the process of classification. Sexual and rapid processing division, the remaining part of the data is tested according to the appropriate rules, and if the data processing requirements are met, then the inductive division is performed. In general, the process of summarizing in data mining includes the result eigenanalysis, the actual number set, the search attribute, the algorithm trade-off, the induction operation, the result visualization processing, the test

set screening, the induction verification principle, the input induction theory and so on. The relevant rules can accurately and detailedly describe the relevant data, and carry out strict analysis and analysis on it. It is necessary to summarize and summarize the specific things in the process of using the law, and then obtain the common mode and then describe it according to the same attribute. Usually the association method is directly applied to the database, which can record the relevant data of each thing as a whole, and improve and improve the performance of the running system based on this. The rough set mainly studies the knowledge of fuzzy and non-deterministic. It belongs to a kind of mathematical tool. The rough set method in data mining does not need to know the information of the corresponding data. It is easier to control the algorithm in the calculation stage; and the rough set can be run. Enrich system functions, which can determine the anomalies in the data, ignore the noise disturbance in the knowledge discovery process, and use the table to find the law of the data in the shortest time to obtain the decision table and improve the convenience of information query. In addition, some uncertainties will appear in some rules. Rough sets will analyze and deal with such uncertain information, which can greatly improve the efficiency of data mining.

3. The Basic Steps of Data Mining

The basic steps of applying data mining in file management include three steps: data collection, data mining and data evaluation.

3.1 Data Collection.

As the name implies, data collection is to select the appropriate archive data to determine the mining target. Different mining results will produce different mining results. If the mining target is determined to be unreasonable, it may lead to mining failure. Therefore, it is necessary to determine the suitable result according to the selected target. The mining plan identifies the issues that need to be addressed first. According to the mining needs, the file information is organically integrated according to different categories and attributes, and then stored in the system database, and the existing relationships between the information resources are analyzed, or the behavior characteristics of the data information are predicted. Data preprocessing needs to eliminate duplicate data, eliminate noise, and convert data types after processing missing data. Data collection transforms data by means of data generalization, smooth aggregation, etc., such as classifying files.

3.2 Data Mining.

Two factors will have an impact on the data mining effect. One is the choice of mining methods, and the other is the quality and scale of data information when mining. Only two elements can obtain the ideal mining results. Mining implementation requires constant feedback, correction, and circulation. Selecting appropriate data analysis tools to solve certain problems and constructing mining models requires continuous optimization, improvement, and calibration of the mining models to ensure that the models are at optimal values. Modeling results need to apply a variety of techniques to compare, obtain appropriate parameters and input variables, pay attention to the choice of variables can not be too much, otherwise the relationship between the data attributes will be affected, the understanding of the results is very unfavorable, if the variables have been selected. If you don't fully understand the relevant attribute characteristics, too little information will be missed. Therefore, when selecting variables, you must ensure that the data quality is good, and there should be no strong correlation between variables.

3.3 Data Evaluation.

After data mining, the system will obtain a series of mining results and model information described in the side. It is necessary to analyze and verify the mining results, and finally determine a reasonable, complete and valuable knowledge system. However, sometimes there may be no meaningful structure, the data collection is unreasonable, the conversion operation is improper, and the result cannot be inconsistent with the mining target. At this time, it is necessary to perform

cyclic data mining, and even to re-determine the mining target. After the data is evaluated, the system language is visualized and converted into a form that is easy for the user to understand.

4. The Specific Application of Data Mining Technology in the File Information Management System

Classification is the core of data mining technology. Its main function is to facilitate the use of data for future prediction and decision-making of information behavior, to mine potential similar properties of target information data, to construct models, and to classify the data to be detected according to a specific method. The classification program mainly analyzes the target data, applies the relevant algorithm to generate the model for each class of the characteristics to be represented by the data, and the decision tree algorithm is the classification method which is the most widely used rapid classification method. The decision tree behaves like a tree trunk shape and is a tree structure of a multi-fork tree. The topmost layer of the tree is the root node, the non-leaf node of the tree is the middle layer, and the leaf node is the final layer of the tree, that is, the class to be divided, and the attributes of the test target are expressed by the branches. In the actual analysis, the top level of the tree should be calculated first, and then tested down to the level until it runs to the leaf node. The final class is the object class. The application decision tree algorithm can not prepare data, and can deal with data type attributes and regular type attributes at the same time. Its measurement standard consists of information entropy and information gain degree, so that it can realize short-time, large-scale data information processing, and finally achieve data classification. The ID3 algorithm will use the inspection category to generate key points and mark the key points. The category data will form a sample of key points. It can be seen that the algorithm can collect the data together, when not used.

The main steps of data preprocessing include three steps: classification of file information, grouping of file information, and data processing of file information confidentiality. The classification of archive information is mainly classified according to the classification criteria of file management, and is presented in the form of primary category, secondary category, and tertiary category; after classification, the file information is grouped, taking the college file management as an example. The archive information can be grouped according to the organizational structure of the university, including the administrative department, the teaching department, the teaching assistant department, etc., and the attributes of different groups are determined according to the division of labor of each department. After determining the grouping of file information, the data is processed according to the confidentiality degree of the file information. First, the original data is sorted, the data is elaborated, the answer is obtained, and the unearthed result is stored in the file information of each department, and the file information presentation form includes Text, pictures, video information, etc., after the information is combined to determine the data, generate graphics. After the original data is sorted out, it is preprocessed to generate valid data for system analysis, eliminating data impurities or noise, so as not to affect the accuracy of data analysis. For example, some college departments lack file information. In this case, the data of the group will usually be initially zero, and the data that does not exist will need to be filled manually. Finally, the data can be converted. The result of the decision tree algorithm is characterized by irregularity. Correspondingly, the data should be irregularly processed. According to the importance and influence of the file information, the confidentiality is divided into top secret and confidential. The three levels of secrets are still based on the management of college archives. The accessors of top secret files are limited to school-level leaders. The inquirers of confidential files can be relaxed to middle-level managers in the school. The secret files are extended to all teaching in the school.

The ID3 algorithm mainly classifies the results of classification and estimation models or estimates and collects unknown results. In the specific implementation process, the data of the archives information department is processed regularly, and the confidentiality degree is classified according to the content of the archive information. It is represented by the form of matrix $A(a_{ij})_{m \times n}$, where m is the number of departments and n is the number of department groups; then the department is classified, and the matching functions between departments are divided into n

department groups, using ID3 theory The matrix A function data is calculated, and the most reasonable data in the calculation result generates key points, and the calculation result is identified to confirm whether they belong to the same type. If the result does not belong to the same type, the matrix A is re-combined to process the data of each function; If the result belongs to the same type, the experimental result is represented by the IF-THEN statement according to the generated decision tree, the data is processed and the decision tree is generated, and the category uses the information gain as the reference standard, and the higher the numerical value of the information gain is used to achieve the calculation target. , set the category with the largest actual result as a key point, determine its properties, and locate the branch Attributes, each branch attribute node then calculates the attribute information gain of each node. Each level is repeated according to the above procedure, and finally a decision tree model is created.

5. Conclusion

In short, with the continuous development of computer information technology, the application of computer mining technology in information processing is more and more extensive. The application of data mining technology in file information management can not only improve the efficiency of file information management, but also improve the convenience of file storage. Sex, and deepen the connotation of file information management, greatly improving the effect of file information application. Of course, in practical applications, data mining technology should be applied reasonably and accurately, and the mode of file management should be innovated while improving the quality of file information management.

References

- [1] Zeng Xuefeng. Development of Computer Data Mining Technology and Its Application in Archives Information Management[J].Science & Technology Innovation and Application, 2016 (9):285.
- [2] Wang Xiaoyan. Application of Data Mining Technology in Archives Information Management [J]. Lantai World, 2014 (23): 25-26.
- [3] Han Jiyi. Construction of Archives Information Management Platform of University Library Based on Data Mining Technology [J].Shanxi Archives, 2015(6):61-63.
- [4] Ha Liyuan. Construction of University Library Archives Information Management Platform Based on Data Mining Technology [J].Shanxi Archives, 2016(5):105-107.
- [5] Blue Sky, Liu Jian, Wang Lancheng. Application of Data Mining Technology in the Field of Information Management [J].Information Research, 2011(2):365-366
- [6] Fang Hongju. The Application of Data Mining Technology in Archive Personalized Service [J].Sci-Tech Information Development & Economy, 2011(10):226-227