

Research and Application of Decision Tree Algorithm and B+Tree Structure on University Smart Employment Platform

Wen Wang, Zhiyu Yang

Guangdong University of Science and Technology, China

Abstract: With the increasing number of graduates and the increasing requirements for employment quality, the traditional information system technology has been difficult to promote the employment of graduates, and its role is mostly limited to the management of graduate information. In view of this situation, this paper will focus on the topic of building a new smart employment system to help college students better obtain employment, analyze the necessity of building a college smart employment platform and the application of decision tree algorithm and B+tree structure in building a college smart employment platform.

Keywords: Graduates; Smart employment platform; Decision tree algorithm; B+tree structure.

1. Current application status and problems of smart employment platform in domestic universities

(1) The practical application of the current graduate employment information system

Nowadays, many colleges and universities begin to have their own information management systems. For colleges and universities, using the information management system to uniformly manage the graduate data will be conducive to the standardized management and search of data, and also provide sufficient conditions for the application of data mining technology.

However, due to the influence of comprehensive factors, most universities still do not have a mature graduate information management system. Through investigation, we found that most of the existing management systems in many colleges and universities have problems such as information redundancy, disordered data classification, low query efficiency, and low utilization of existing data and waste of information system resources. In combination with the actual situation, it can be seen that these problems are still difficult to solve in a short time.

(2) Problems in the future development of current graduate employment information system

To sum up, for most domestic colleges and universities, it is necessary to explore a new way to improve the operational capacity of the information management system. Based on this problem, data mining and other related technologies can be integrated into the graduate employment information system to establish a new smart employment platform.

(3) Function introduction of the new smart employment platform

The general function of the smart employment platform is to comprehensively improve the traditional employment platform. In addition to improving the operation performance of the system itself, it can also integrate online information data and offline market resources, fully enable college employment with Internet technology, realize online employment behavior, and finally build a convenient and efficient "information bridge" between college students'

employment and enterprise recruitment.

Users can create accounts on the smart employment platform, and the platform will save users' information for logging in at any time. Users can use the platform to record their employment related information, such as rewards and punishments, students' obtaining certificates, graduation destinations, employment preferences and other information conducive to the algorithm to find matching information. At the same time, the system can also collect the recruitment information and recruitment requirements of various enterprises with the help of the network and other resources, and then match and search with the existing graduate information, so as to improve the adaptability of students and enterprises as much as possible. On this basis, the platform will also push the recruitment information and market information that meet their input information to users, so as to ensure that they can have information sources to choose more suitable jobs at any time. If this technology is fully promoted, the employment rate and quality of graduates will be greatly improved.

2. Analysis of the actual structure of the smart employment platform

Data mining needs to combine technologies in multiple fields to play a powerful role, and technologies in each field work on different architecture layers. Therefore, this process requires coordination and cooperation of multiple hierarchies to ensure that user needs can be correctly judged, and matching information can be found in the database according to needs, and then the information can be sent back to users. Based on this problem, a mature data mining system should include the following three layers from the inside out: DBMS layer, central algorithm layer and client layer.

As a complete system, it is natural to have user-oriented ports, so the outermost layer of the entire system is the client layer first. The central algorithm layer interfaces with the client and back-end DBMS. The architecture layer uses decision tree algorithms to help find information, which plays a pivotal role in communicating users and databases.

In addition, the smart employment platform can also collect users' interests and expectations, and intelligently push

relevant recruitment information to users. At the same time, users can also evaluate the recommendation results and search results to test the accuracy of the smart employment system, so as to facilitate the subsequent update and improvement of the system.

3. Application of decision tree algorithm to central algorithm layer

(1) Overview of Data Mining Technology

Data mining, as the name implies, is to mine useful information from massive data. In various databases, there is often a large amount of information, which is integrated through DBMS management and stored in the database. However, because of its huge information retention and increment, if read one by one, it will not only be inefficient, but also produce a lot of redundant information. Data mining can greatly improve the pertinence of information processing, and improve the technical support of data analysis with different needs.

(2) Brief Description of Decision Tree

Classification technology is a systematic method to establish classification models based on input data sets, and is a common technology in the process of data mining. The classification technology is mainly used to divide the observed values uniformly according to the established categories. A common algorithm in classification technology is the decision tree algorithm, so the central algorithm layer will use the decision tree algorithm as the core algorithm to achieve the idea.

(3) Example of Decision Tree Generation

For a data set, if data mining is adopted, 70% of the data will be used for training and 30% for testing. From this, a relational model can be generated, which can be used to predict tag results when the attributes of related data are known. If the difference between the predicted results and the actual results is not large, it indicates that the model is a good model, otherwise the model still needs to be improved.

Gini coefficient is taken as the index to measure the rationality of each attribute. Starting from the root node, it is necessary to calculate the Gini coefficient of each attribute, select the attribute with the lowest Gini coefficient as the root node, and use the Gini coefficient formula as a reference.

$$Gini = 1 - \sum_{K=1}^n [p(1) + p(2) + \dots + p(i)]^2$$

Figure 1. Gini coefficient formula

After the principle of decision tree and its generation process are known, the design idea can be incorporated into the algorithm development. In the development and application of algorithms, there is another important process that cannot be ignored, that is, data filtering. In the initially acquired data, there will be a lot of noise data, redundant data, and some data may even be disordered and missing. If such data is directly used in data mining, the quality of extracted data will be reduced. Therefore, these problematic data should be processed in advance with the help of data mining algorithms to improve the accuracy of subsequent information acquisition.

Therefore, the application of decision tree algorithm to the intelligent employment system will significantly improve the user experience compared with the traditional employment

information management system. Graduates will get more accurate information after decision tree algorithm matching and screening, which is conducive to shortening graduates' waiting period and reducing their employment decision-making costs. At the same time, the decision tree algorithm can also be used to predict and obtain the latest situation of the market demand for talents.

4. Application of B+Tree Structure

In addition to the central algorithm layer, the smart employment platform also has a very important layer, namely the bottom DBMS layer. If the decision tree structure adopted by the central algorithm layer can improve the accuracy of data mining, the ability of DBMS to quickly find information will directly determine the information matching speed and information quality of the entire intelligent employment platform. At this level, because we often have to face a huge amount of data, and also need to respond to the search instructions of the algorithm layer in the shortest possible time, so it is necessary to optimize the underlying database. As a result, the B+tree structure can be used to optimize the internal structure of the underlying database to reduce the number of IO and improve the speed of searching information.

The structure of B+tree can help the paging management of all levels of pages in the database, and help the database to obtain the required data efficiently. It can directly find the lower limit of the range by using the dichotomy search method, and then traverse the list generated between leaf nodes in order until the upper limit of the range is found. The intermediate node with k subtrees in the B+tree contains k elements. Each element does not store data but is only used for indexing, that is, the number of elements in the node corresponds to the number of subtrees. Each leaf node contains all the element information, and points to the specific records of these elements with the help of pointers. The leaf nodes themselves are sorted from small to large depending on the chain structure. All intermediate node elements will exist in the child nodes, which are the largest or smallest elements in the child nodes.

The paging hierarchical structure of B+tree can also reduce the coupling degree of each node. Traditional database structure mostly adopts sequential structure, and lacks paging hierarchical structure. The coupling between data modules is high. Once a part of the database fails, it will most likely lead to the collapse of the entire system, thus affecting the user experience and increasing the maintenance cost of the platform. After using the B+tree structure optimization, even if a part of the database fails, other parts can still be used normally after the paging management is implemented, which can effectively avoid the collapse of the entire database server.

5. Summary

Through the analysis of the overall situation of the current domestic university information management system resource application, it is found that there are some problems such as immature information management system and insufficient utilization of information. In addition to such problems, there are some system performance problems such as the backward information screening and matching function of the traditional employment management system, the system is easy to collapse, and the information query efficiency is low. Based on these problems, we can build a smart employment system consisting of the underlying

database layer, the central algorithm layer and the client layer. The central algorithm layer uses the decision tree principle to mine the market recruitment information and the graduate information to match, so as to promote the employment demand of graduates to match the demand of enterprises for talents. The bottom database layer uses the B+tree paging hierarchical structure to realize the rational management of the database, reduce the system's IO times, improve the database's operating efficiency, and reduce the possibility of system crash. Through the combination of various technologies, the goal of improving the traditional employment system of college graduates and promoting high-quality employment of college graduates can be finally achieved.

Acknowledgment

This work is supported by the 2022 school level scientific research project (natural science) of Guangdong University of Science and Technology (XJ2022004001).

References

- [1] Yang Zhiyu, Wu Shiyu, Wang Zengqi, Ding Guoyong Analysis of applied technologies in the field of domestic education data mining in the past decade [J], China Education Informatization, 2022,28 (04): 62-69.
- [2] Wang Jiejie, Tong Dongfeng. Analysis on the Path of Building a Smart Employment Platform in Colleges and Universities [J]. Public Relations World, 2022, (13): 90-92.
- [3] He Jin. Research on the Application of Data Mining Technology in the Employment Information Management System of College Students [J]. Heilongjiang Human Resources and Social Security, 2022, (13): 79-81.
- [4] Wang Kai, Wang Na, Guo Lixia. Research on the Application of Data Mining Technology in the Employment of College Graduates [J] Modern Information Technology, 2020, 4 (08): 83-85+89.