Intelligent Query Method of Vocational Colleges Database Based on Reinforcement Learning

Jing Liu

Liaoning vocational university of technology, Jinzhou, China

Email:liujingjing@163.com

Abstract. Due to the low adaptability to the characteristics of the database of vocational colleges, it is difficult to guarantee the effectiveness of the intelligent query results of the database. Therefore, the research on the intelligent query method of the database of vocational colleges based on reinforcement learning is proposed. Combined with the core characteristics of the database of vocational colleges, such as large amount of data, data redundancy, data heterogeneity, and frequent data updates, the core idea of improving the final query effect of the design method is to comprehensively consider the factors such as data quality, query statements, retrieval algorithms and models, threshold settings, and data distribution. The enhanced learning mechanism is introduced to realize the comprehensive learning of the data distribution characteristics in the database of vocational colleges and universities, and the optimal parameters that meet the characteristics of the database of vocational colleges and universities are taken as the query output results. In the test results, the F1 value of the design query method is always above 0.94 in different scenarios, which has obvious advantages over the query method of the control group.

Keywords: Reinforcement learning; Database of vocational colleges; Intelligent query; Core characteristics; Core idea; Distribution characteristics; After comprehensive study.

1 Introduction

At this stage, the database query of vocational colleges has become one of the most important auxiliary means in college management. First, in terms of education management, schools need to manage a large number of student information, faculty information, curriculum information, performance information, etc. With the help of database intelligent query technology, these data can be efficiently stored and queried[1-3]. Secondly, in terms of student services, as an important responsibility of the school, including student file management, course selection management, scholarship management, etc., database intelligent query can facilitate a large number of data query operations. In addition, in terms of faculty management, faculty information management, salary management, teaching task allocation, etc.[4-5] also need to use database intelligent query technology to quickly and accurately query data. Finally, in terms of

M. Yu et al. (eds.), *Proceedings of the 2024 5th International Conference on Big Data and Informatization Education (ICBDIE 2024)*, Advances in Intelligent Systems Research 182,
https://doi.org/10.2991/978-94-6463-417-4_51

decision support, school leaders need to make decisions based on a large amount of data, such as making teaching plans, evaluating education quality, etc. Database intelligent query technology can quickly query and analyze data. With the continuous advancement of school informatization construction, data between various departments may be inconsistent. This requires database intelligent query technology to strengthen data integration [6-7] to achieve unified management and query of data. To sum up, vocational colleges and universities have a wide range of application needs for database query, while there is also some room for optimization. In order to improve the efficiency and security of the database, it is necessary to design and manage the database scientifically[8-10]. Among them, the database query method based on Delphi is based on the component oriented programming language Delphi. From the perspective of database performance, it designs the Select statement query, fuzzy query and parametric query, which greatly improves the query effect, but the efficiency is low; The database query method based on multithreading programming is also based on Delphi language, which improves the efficiency of query, but there is room for further improvement in precision[11-12].

On the basis of the above, this paper proposes the research on the intelligent query method of the database of vocational colleges based on reinforcement learning, the core idea of improving the design method's final query effect is based on the core characteristics of vocational undergraduate college databases. Based on reinforcement learning, comprehensive learning of data distribution characteristics in vocational college databases is achieved, and the optimal parameters that meet the characteristics of vocational college databases are used as query output results. By designing a comparative testing environment to analyze and validate the performance of the designed query methods, we can ensure the effectiveness of the intelligent query results of the database.

2 Design of intelligent query method for database of vocational colleges

2.1 Analysis of database characteristics of vocational colleges

In order to realize the effective design of the intelligent query method for the database of vocational colleges, it is necessary to fully understand and consider the characteristics of the database of vocational colleges and the main factors that affect the final query results [13-14]. As far as the structural characteristics and data distribution characteristics of the database of vocational colleges and universities are concerned, they will vary according to the specific conditions and actual needs of the schools. However, in general, the database of vocational colleges and universities will usually use relational databases, such as MySQL, Oracle, etc. These databases have a rigorous data structure, and can well organize and manage various data such as student information, faculty information, curriculum information[15-16]. In relational databases, data is stored in the form of tables. Each table consists of rows and columns. Each row represents a record and each column represents an attribute [17-18]. On this basis, the data in the

database of vocational colleges and universities show more prominent characteristics in the specific distribution. This paper summarizes them as four points shown in Table 1.

Table 1. Distribution characteristics of database data of vocational colleges

Number	Characteristic	Form of expression
1	Large data volume	Due to the large amount of student information, faculty information, and course information that schools need to process, the data volume in databases is usually large.
2	Data redundancy	Due to repeated data entry and backup, there may be some data redundancy in the database.
3	Data heterogeneity	Data from different departments may have certain heterogeneity, such as student information, including basic information, grade information, award information, etc. Different data structures may exist between different information.
4	Frequent data updates	Due to the school's need to constantly update data, such as student information and course information, the data in the database may be updated frequently.

According to the analysis of the characteristics of the data in the database of vocational colleges and universities in the specific distribution in Table 1, we can see that the core characteristics of the database are large data volume, data redundancy, data heterogeneity and frequent data updates[19-21]. In this context, the main factors affecting the query accuracy of the database of vocational colleges and universities analyzed in this paper include the five aspects shown in Table 2.

Table 2. Analysis results of factors influencing the query accuracy of the database of vocational colleges

Number	Influence factor	Form of expression
1	Data quality	If the data quality in the database is low or there is noise, missing, or error, it may lead to missed or false detections, thereby affecting query accuracy.
2	Query statements	The expressive power and accuracy of query statements directly affect the quality of search results. If the query statement is inaccurate or incomplete, it may result in missing relevant documents or returning a large number of unrelated documents, thereby affecting query accuracy.
3	Retrieval algorithms and models	Different retrieval algorithms and models have different impacts on query accuracy. Some algorithms may focus more on improving recall, while others may focus more on improving precision.
4	Threshold setting	In some retrieval systems, it is necessary to set a threshold to determine when a document is considered relevant. Lower thresholds may increase recall, but may result in lower precision, while higher thresholds may increase precision, but may result in lower recall.
5	Data distribution	The uneven distribution of data can also affect query accuracy. In some cases, if the data is too concentrated or scattered, it may lead to inaccurate query results.

Data quality: Use data cleaning and preprocessing techniques to eliminate noise, correct errors, fill in missing values, and standardize data. These technologies can improve the accuracy, consistency, and completeness of data, thereby providing more reliable and consistent results for queries.

Query statements: Using natural language processing (NLP) technology to parse, process, and understand user query statements. NLP technology can help identify keywords for queries, extract entities and attributes, expand query scope, and more accurately match and retrieve relevant data.

Retrieve algorithms and models: Use advanced retrieval algorithms to improve query performance. This includes TF-IDF (Word Frequency Inverse Document Frequency) algorithm, BM25 (Best Matching 25) algorithm, and vector space model. These algorithms determine the most relevant results by calculating the similarity between queries and documents, and sort them based on their relevance.

Threshold setting: Determine which results are considered relevant by setting appropriate thresholds. The threshold can be based on specific evaluation indicators, such as correlation score, matching degree, etc. This can filter out results that are not related to the query, improving the accuracy and efficiency of the query.

Data distribution: Using distributed computing and data distribution technology to distribute data across multiple nodes to improve query performance. By distributing data to multiple nodes and running parallel queries, it is possible to quickly respond and provide high throughput query results.

To sum up, the structural characteristics and data distribution characteristics of the database of vocational colleges and universities have certain characteristics. At the same time, there are many factors that affect the query accuracy. In order to improve the query precision, when designing the specific query method [22-23], the core idea of improving the final query effect of the design method is to comprehensively consider the factors such as data quality, query statements, retrieval algorithms and models, threshold settings, and data distribution.

2.2 Based on reinforcement learning vocational colleges database intelligent query

Reinforcement learning is a machine learning method aimed at learning optimal behavior strategies through interaction with the environment. In terms of optimizing query results, reinforcement learning can be used to improve search engine ranking algorithms to provide more relevant and accurate search results. According to the characteristics of the database of vocational undergraduate colleges, different states can be defined. Define actions such as sorting and filtering criteria for query results. Design a reward function based on user feedback and goals to evaluate the quality of intelligent query results. Based on reinforcement learning algorithms to train the model, the model can learn to select the best action in different states to maximize the reward function. Apply the trained model to the actual database of vocational undergraduate colleges and continuously update the model through interaction with users. When users perform queries, the model can select the best action based on the current state and adjust the model parameters based on user feedback.

Combined with the above-mentioned characteristics of database in vocational colleges and the core idea of improving the final query effect of intelligent query method, this paper introduces reinforcement learning mechanism, and uses it to realize comprehensive learning of data distribution characteristics in vocational colleges' database [24-25]. Among them, the specific learning process is shown in Figure 1.

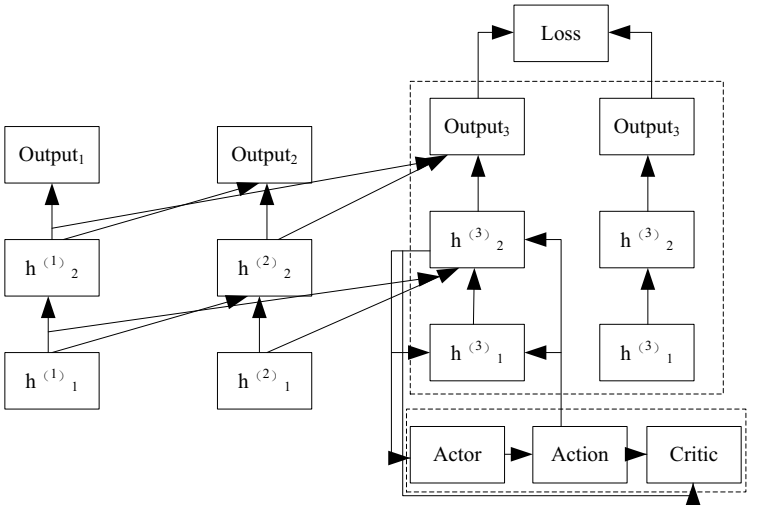

Fig. 1. Comprehensive learning method based on reinforcement learning for database data distribution characteristics of vocational colleges.

According to the way shown in Figure 1, the characteristic information of database data distribution in vocational colleges is taken as the characteristic parameter of deep learning network in this paper, and expressed in the form of tensor, specifically as follows formula (1):

$$T = (ty, h, \omega, flps, stride, k, r) \tag{1}$$

Among them, T represents the characteristic tensor parameters of the data distribution of the vocational undergraduate school database, ty represents the type of the current hierarchy, h represents the characteristic dimension parameters of database data distribution input at each layer, ω represents the output of each layer for the database data distribution characteristic dimension parameter , $flps$ represents the amount of computation, that is, the complexity of the reinforcement learning network, $stride$ represents the step size of each layer, k represents the neuron size of the reinforcement learning network, r represents the ratio of the size of the network output channel after pruning to the original network output channel.

On this basis, the feature tensor shown in Formula (1) can be used to constrain the change of the output learning results of the current network structure level and form a state space structure in the reinforcement learning problem. At this time, the compres-

sion rate of each layer of the network output is the absolute value of the parameter gradient of each layer. In this paper, when sorting the parameters of each layer structure in the reinforcement learning network, it is also based on the absolute value of the corresponding parameter gradient, and the parameters with the highest absolute value of the parameter gradient are reserved, that is, the parameters with high importance, so as to realize the comprehensive learning of the data distribution characteristics of the database.

On this basis, when fusing the features of the query task, this paper uses the attention unit to encode and fuse the information in the data, and with the help of the attention weight, fuses the input features of the corresponding query task to form a new output form, in which the purpose of inputting the query task into the attention layer is to retain the supporting information of the previous query task knowledge for the current query task output, and the output result of the attention unit can be expressed as follows formula (2):

$$A_m = \tanh(W_m T_m + b_m) \tag{2}$$

Among them, A_m indicate that output result of the attention unit to the current query task on the premise that there are m tasks before, W_m indicate that weight of the attention parameter, T_m represents the previous m specific query task information, b_m represents the absolute value of the parameter gradient.

Among them, the distribution of attention parameter weights under the influence of data characteristics of vocational undergraduate colleges' databases can be expressed as follows formula (3):

$$W_m = \frac{eT_m * m}{\sum eT_m * m} \tag{3}$$

At this time, according to the overall attention parameter weight distribution matrix, the output query information can be expressed as follows formula (4):

$$S = \max \sum W_m H_m \tag{4}$$

Among them, S represents the output query result, H_m represents the shared characteristic parameters of database data in vocational colleges.

According to the way shown above, the design of intelligent query method for database of vocational colleges is realized.

3 Application testing

3.1 Test preparation

In the practical application effect stage of the intelligent database query method based on reinforcement learning designed in this paper, a comparative test is carried out. Among them, the query methods used in the control group are the database query method based on Delphi and the database query method based on multi-thread programming. The specific test environment is designed. In the independent test environment, MySQL database is selected as the storage engine, and Python programming language is used to write intelligent query algorithms. A certain scale data set is extracted from the actual vocational college database for testing to ensure that the data distribution characteristics are consistent with the actual situation.

On this basis, the following three test scenarios are designed for the intelligent query method of database in vocational colleges as shown in Table 3.

Table 3. Test Scenario Design

Test scenario number	Testing method	Specific settings
Scenario 1	Keyword based query	Perform keyword queries on data tables such as student information, course information, and faculty information, including fields such as student name, course name, and teacher name.
Scenario 2	Query based on multi table association	In multi table association queries, we need to perform query operations based on the relationships between multiple data tables. For example, querying a student's course selection and their corresponding teacher information.
Scenario 3	Time series based queries	In temporal queries, we need to sort and filter data based on temporal order. For example, to query the changes in students' grades within the past month.

In the testing process, we use F1 score (an evaluation index that comprehensively considers precision and recall) to evaluate the performance of different query methods.

In the above test environment, the query performance of different methods is tested respectively.

3.2 Test results and analysis

The specific test results in different test scenarios are shown in Figure 2.

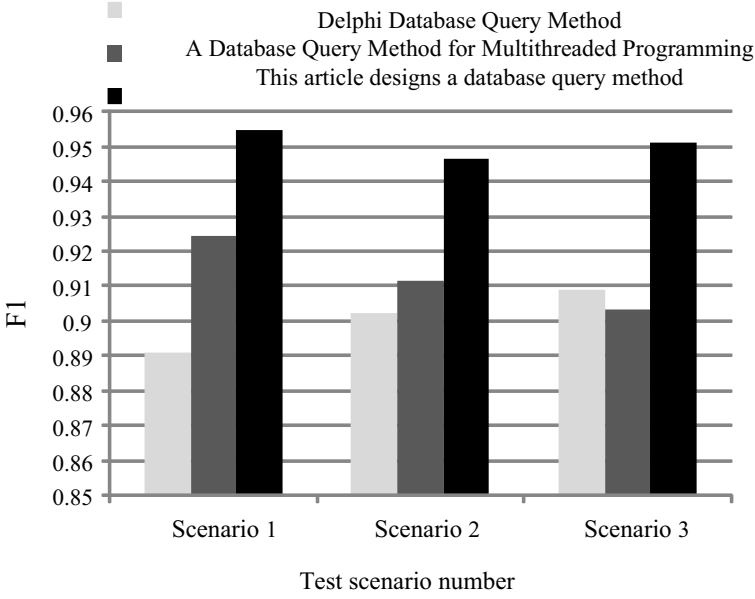

Fig. 2. Comparison chart of test results of different methods

Combined with the test results shown in Figure 2, it can be seen that the F1 value of the query method designed in this paper is always above 0.94 in different scenarios, which has obvious advantages compared with the query method of the control group. It shows that it can effectively query information.

4 Conclusion

By analyzing the actual development level of database query in vocational colleges at this stage, we can see that there is also some room for optimization. First of all, in terms of database design, in some cases, the design of the database may not be reasonable enough, resulting in low query efficiency. In some query operations, unreasonable query statements or insufficient indexes may lead to slow query speed. Moreover, considering the vital position of data security, establishing a perfect data security guarantee mechanism is also an important basis to ensure the confidentiality and integrity of data. This paper puts forward the research on the intelligent query method of database in vocational colleges based on reinforcement learning, which effectively improves the reliability of the query effect.

References

1. Hatcher W G, Qian C, Gao W, et al. Towards efficient and intelligent internet of things search engine[J]. IEEE Access, 2021, 9: 15778-15795.
2. Yan H, Li X, Li H, et al. Monitoring-based differential privacy mechanism against query flooding-based model extraction attack[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 19(4): 2680-2694.
3. Suryotrisongko H, Musashi Y, Tsuneda A, et al. Robust botnet DGA detection: Blending XAI and OSINT for cyber threat intelligence sharing[J]. IEEE Access, 2022, 10: 34613-34624.
4. Liu N, Shapira P, Yue X. Tracking developments in artificial intelligence research: constructing and applying a new search strategy[J]. Scientometrics, 2021, 126(4): 3153-3192.
5. Malik A, De Silva M T T, Budhwar P, et al. Elevating talents' experience through innovative artificial intelligence-mediated knowledge sharing: Evidence from an IT-multinational enterprise[J]. Journal of International Management, 2021, 27(4): 100871.
6. Laeeq K, Memon Z A. Scavenge: An intelligent multi-agent based voice-enabled virtual assistant for LMS[J]. Interactive Learning Environments, 2021, 29(6): 954-972.
7. Guan Y, Lu R, Zheng Y, et al. Toward privacy-preserving cybertwin-based spatiotemporal keyword query for ITS in 6G era[J]. IEEE Internet of Things Journal, 2021, 8(22): 16243-16255.
8. Balaji B S, Balakrishnan S, Venkatachalam K, et al. Automated query classification based web service similarity technique using machine learning[J]. Journal of Ambient Intelligence and Humanized Computing, 2021, 12: 6169-6180.
9. Wolff J, Pauling J, Keck A, et al. Success factors of artificial intelligence implementation in healthcare[J]. Frontiers in digital health, 2021, 3: 594971.
10. Ram J, Zhang C. Examining the role of social media analytics in providing competitive intelligence: The impacts and limitations[J]. Journal of Global Information Management (JGIM), 2021, 29(6): 1-18.
11. Hoffman S C, Chenthamarakshan V, Wadhawan K, et al. Optimizing molecules using efficient queries from property evaluations[J]. Nature Machine Intelligence, 2022, 4(1): 21-31.
12. Dong Y, Cheng S, Pang T, et al. Query-efficient black-box adversarial attacks guided by a transfer-based prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(12): 9536-9548.
13. Evangelista J R G, Sassi R J, Romero M, et al. Systematic literature review to investigate the application of open source intelligence (OSINT) with artificial intelligence[J]. Journal of Applied Security Research, 2021, 16(3): 345-369.
14. Loureiro S M C, Guerreiro J, Tussyadiah I. Artificial intelligence in business: State of the art and future research agenda[J]. Journal of business research, 2021, 129: 911-926.
15. Gao P, Li J, Liu S. An introduction to key technology in artificial intelligence and big data driven e-learning and e-education[J]. Mobile Networks and Applications, 2021, 26(5): 2123-2126.
16. Osamy W, Khedr A M, Salim A, et al. Recent studies utilizing artificial intelligence techniques for solving data collection, aggregation and dissemination challenges in wireless sensor networks: a review[J]. Electronics, 2022, 11(3): 313.
17. Moor M, Banerjee O, Abad Z S H, et al. Foundation models for generalist medical artificial intelligence[J]. Nature, 2023, 616(7956): 259-265.

18. Gupta K, Jiwani N. A systematic Overview of Fundamentals and Methods of Business Intelligence[J]. International Journal of Sustainable Development in Computing Science, 2021, 3(3): 31-46.
19. Wang T, Yang Q, Shen X, et al. A privacy-enhanced retrieval technology for the cloud-assisted internet of things[J]. IEEE transactions on industrial informatics, 2021, 18(7): 4981-4989.
20. Rasheed I, Banka H, Khan H M. Pseudo-relevance feedback based query expansion using boosting algorithm[J]. Artificial Intelligence Review, 2021, 54(8): 6101-6124.
21. Adetayo A J. Artificial intelligence chatbots in academic libraries: the rise of ChatGPT[J]. Library Hi Tech News, 2023, 40(3): 18-21.
22. Liu Y, Yu J, Fan J, et al. Achieving privacy-preserving DSSE for intelligent IoT healthcare system[J]. IEEE Transactions on Industrial Informatics, 2021, 18(3): 2010-2020.
23. Pavlik J V. Collaborating with ChatGPT: Considering the implications of generative artificial intelligence for journalism and media education[J]. Journalism & Mass Communication Educator, 2023, 78(1): 84-93.
24. Rajendran S, Mathivanan S K, Jayagopal P, et al. Emphasizing privacy and security of edge intelligence with machine learning for healthcare[J]. International Journal of Intelligent Computing and Cybernetics, 2021, 15(1): 92-109.
25. Jiang S, Gao Y, Xv Y L. Redundant Data Query of Spatial Database Location Based on Fruit Fly Algorithm [J]. Computer Simulation, 2022(008):039.