Optimization of Data Query Method Based on Fuzzy Theory

Yunwei Li[1(✉)] and Lei Ma[2]

[1] Beijing Youth Politics College, Beijing 100102, China
liyunwei@bjypc.edu.cn
[2] Beijing Polytechnic, Beijing 100016, China

Abstract. In order to achieve the research goal of fast and accurate query of massive complex data, this study proposes a data query optimization method based on fuzzy theory. Firstly, the characteristics of the data to be queried are identified combined with the fuzzy theory, and the characteristics are classified. Then, the data management model is constructed to optimize the data query management process. The experimental results show that this method can effectively ensure the accuracy and comprehensiveness of massive data query, and prove that it can fully meet the research requirements.

Keywords: Fuzzy theory · Data query · Data management · Feature recognition · Feature classification

1 Introduction

There are a lot of fuzzy or uncertain information in relational database, and the traditional data query method has a poor effect on this kind of information, mainly manifested in the inaccurate query results. In this case, relevant scholars have proposed a probabilistic data query method based on sensor detection, and carried out research on nearest neighbor and probabilistic query problems [1]. However, it is worth noting that the probability model is only a theoretical model from the perspective of data processing, and does not give a query model according to the characteristics of the data itself and the actual needs of users [2].

Based on the above analysis, this paper studies the data storage and query distribution mechanism in fuzzy sensor networks, and designs an optimization scheme of data query method based on fuzzy theory. In this study, according to the characteristics of sensor network data, the data storage mechanism and query distribution mechanism are studied. On this basis, the features of the data to be queried are identified by fuzzy theory, and a data management model is established after the features are classified. In the model, a plurality of query areas are divided, and the data characteristics of each area are returned to the base station according to the query route. The base station obtains the final fuzzy theory value through sorting, so as to achieve the data query goal.

W. Fu and L. Yun (Eds.): ADHIP 2022, LNICST 468, pp. 56–67, 2023.
https://doi.org/10.1007/978-3-031-28787-9_5

2 Data Query Method Optimization

2.1 Fuzzy Data Feature Extraction Management

With the development of fuzzy theory and technology, a special intelligent sensor that can accept, generate and process fuzzy signals-fuzzy sensor came into being [3]. Fuzzy sensor, also known as symbol sensor and fuzzy symbol sensor, is an intelligent sensor that outputs the measurement results in the form of natural language symbol description after fuzzy reasoning and integration based on the numerical measurement of classical sensors. The research of fuzzy sensor is a late branch in the application of fuzzy logic technology [4].

With the maturity of fuzzy sensor theory, many fuzzy sensors have been applied, such as distance fuzzy sensor, chroma fuzzy sensor, temperature and humidity fuzzy sensor, comfort fuzzy sensor and so on. In order to obtain the results of fuzzy query, there are often two basic methods: one is to establish a fuzzy database model and modify its database structure by adding fields containing fuzzy attributes. The second method is to establish an accurate relational database based on fuzzy logic and fuzzy calculation of query conditions by constructing membership functions or fuzzy expansion of SQL statements, Transform it into a fuzzy range, and then conduct accurate SL query. The second method is used to define the fuzzy elements in the feature data management database, and through Visual FoxPro 6.0 converts fuzzy query into equivalent SL query statement to realize fuzzy query instance [5].

Addition, deletion and modification are the basic operations of the database. In order to complete these operations, you need to query the corresponding records in the database and locate them quickly. The current basic query of relational database is very easy to implement for precise fields such as “gender”. Part of the information cannot be located correctly due to its fuzziness [6]. Based on fuzzy mathematics and fuzzy set, a SQ query method combined with fuzzy theory is proposed according to the method of reasonably self-defining membership function, and a query example is realized on MicrosoftVisualFoxpro6.0 platform to optimize the feature extraction process of fuzzy data. The specific process is shown in Fig. 1.

It is assumed that all data are deterministic. That is, the data item of each sensor is the most real situation of the detected external environment [7]. In the process of query, these exact values are used to return the results. Although the query is satisfied by calculating the true and exact value, this situation will still lead to incorrect query results, as shown in Fig. 2.

It is assumed that the user queries the temperature value of a sensor. It is assumed that the temperature detected by the sensor is very accurate, and the transmission will not cause the imprecision of the data. However, when a user queries the database and finds that the current data are A0 and B0, the ambient temperature may already be A1 and B1, so the query is inaccurate.

Because the point uncertainty model often leads to accurate data and inaccurate results, an interval uncertainty query model is proposed. The data in the database is different from the traditional data. It replaces the original point form in the form of an interval. The uncertain data interval is defined as U. In particular, U_i is a real valued function of a closed interval with respect to t, which is used to limit the value of Ta at

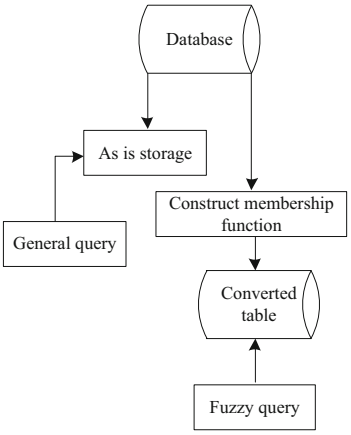

Fig. 1. Fuzzy Data feature extraction process

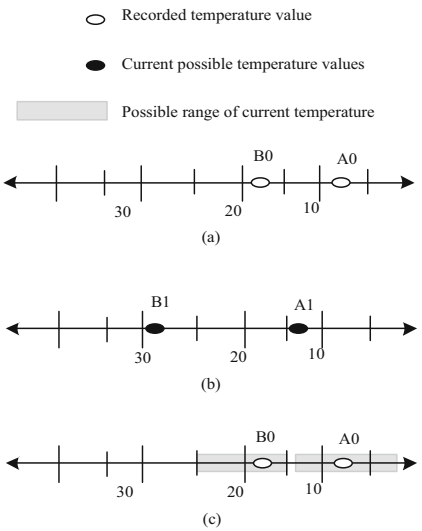

Fig. 2. Example of point query and interval direct query of uncertain data

time t. The model describes the imprecise data in the form of interval, so that U can change linearly according to time until the next update arrives.

2.2 Fuzzy Data Processing Algorithm

Fuzzy data processing is based on the transformation from digital information to symbolic information. In order to realize symbolic measurement, it is necessary to clearly give the corresponding relationship between symbols and numbers to ensure that two different symbols cannot have the same meaning [8]. Based on this, the change principle of fuzzy semantics and fuzzy description is displayed, as shown in Fig. 3.

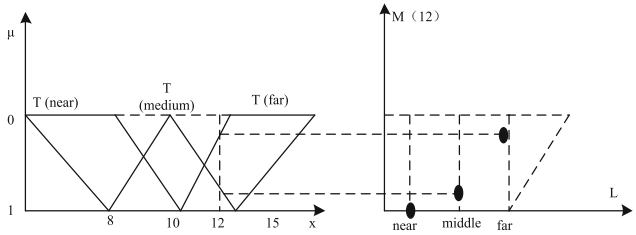

Fig. 3. Fuzzy Semantics and fuzzy description

In this process, there are five ways to query according to various classification methods: according to the query time, it can be divided into query at a certain time, query in a time period and continuous query. According to the value returned by the query, it can be divided into value based query and entity based query [9]. According to aggregation or not, it can be divided into spatial aggregation query, temporal aggregation query and non aggregation query. According to the filtering method, it can be divided into filtering and non filtering. According to the dimension of perceived data, it can be divided into one-dimensional and multi-dimensional queries. Queries based on these classification methods are divided into 28 categories. Although the model in this paper can be extended to all the above 28 queries, in order to facilitate the discussion later in this paper, this paper focuses on two cases, as shown in Table 1.

Table 1. First level query classification method

QUERY CLASS		Fuzzy query
TIME	Point of time	√
	Continuity	
	Time slot	√
ANSWER	Entity	√
	Value	√
AGGREGATION	Spatial aggregation	√
	Time aggregation	
	Non aggregation	√
FILTERING	filter	
	Non filtering	√
DIMENSIONS	Apile	√
	Multi pile	

The answers returned by all these aggregate queries are in the form of fuzzy language and membership degree, which should meet the constraints of fuzzy membership function. Table 2 summarizes the methods of fuzzy query. This table lists the differences between fuzzy query and ordinary sensor query, and finally gives the values given by the query.

Table 2. Classification of fuzzy sensor query methods

Query type	Entity based query	Numeric based query
Gather	FKNN, FENNQ, FEMinP, FEmaxP	FLAvgP, FLSumP, FLMaxP
Non aggregation	FERP	LSingleP
Membership fuzzy query	Sq, [μ1,μu]	Lq,[μ1,μu]
General query	Sq	V

It can be seen that fuzzy query gives more query result information than ordinary query. The query methods in this table are some of the query methods of fuzzy sensor networks. In practical applications, many have been widely used, such as fuzzy theory query. The random walk mechanism is used to search the data query request node in the process of network communication. The approximate nodes and equivalent nodes of the node are searched twice, and the nodes and data with lower load and higher evaluation value are returned. These two are used as the query target nodes and data required by the user. The specific process is as follows: Suppose that the number of service nodes providing user data query in wireless sensor network is d, and all query service nodes form an overlay network c. the set of user data objects on service node n is represented by $b_{|dc_n|}$, and the number of user data objects is $b_{|dc_n||dk_n|}$. For any service node n in wireless sensor overlay network k, further describe the i user data center, The network bandwidth between data centers of different users in B can be expressed as

$$B_i = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{|dc_n|} \\ \vdots & \vdots & \ddots & \vdots \\ b_{|dc_1|} & b_{|dc_2|} & \cdots & b_{|dc_n||dk_n|} \end{pmatrix} \tag{1}$$

Membership function is a basic concept in fuzzy mathematics. Fuzzy set is completely determined by its membership function. When applying fuzzy mathematics to solve practical problems, the determination of membership function is very important. It is the first step to apply fuzzy mathematics to solve practical problems. Considering the fuzziness of entities, it must be incorrect to use binary characteristic function to express it [10]. In order to express these attributes more accurately, so that their value in [0,1] means that the value can be unlimited, the data characteristic membership function is obtained as follows:

$$\mu_A = \left\{ \begin{array}{ll} B_i & x \leq 30 \\ \frac{b_{|dc_n|}}{10}, & 30 < x \leq 40 \\ 1, & 40 < x < 50 \\ \frac{b_{|dc_n||dk_n|}}{5}, & 50 \leq x \leq 55 \\ 0, & x > 55 \end{array} \right\} \tag{2}$$

Different types of user queries have a great impact on the query methods and efficiency of encrypted data. This paper divides user queries into the following two categories:

(1) Attributevalue query;
(2) Attributeopattribute join query.

Among them, attribute is an encrypted attribute of the trusted database, WLE is the query value for attribute, and OP is the query operator, which can be "=", ">", "<", "", and so on. For a series of attribute = value queries, set the ciphertext data segment of the current encryption attribute A as Q_K, a total of k segments, the ciphertext tuple size of each data segment is $M_1, M_2, \cdots, M_I$, the number of times the query value falls in each ciphertext data segment is L and Z respectively, and the total number of correct results returned by each ciphertext data segment is $H_1, H_2, \cdots, H$, then the total false positive rate of the query is as follows:

$$P = \mu_A - \frac{\sum_{i=1}^{k} B_i}{Q_K \sum_{i=1}^{k} LZ + M_i} \tag{3}$$

Adaptive index partition strategy is adopted to dynamically adjust the segmentation of ciphertext with the change of user query and data Firstly, the influence of ciphertext data segmentation on the false positive rate of query is analyzed. Suppose that the current i ciphertext data segment is re divided into two new data segments, each containing the number of ciphertext tuples M_{i1} and M_{i2}, carrying the query service Q_{i1} and Q_{i2}, and the total number of correctly hit query results is H. on the premise that other ciphertext data segments remain unchanged, the change of false positive rate before and after segmentation is as follows:

$$\lambda = \frac{\sum_{t=1}^{k} P + H}{1 + \sum M_{i1} Q_{i1} + M_{i2} Q_{i2}} \tag{4}$$

There is only one data segment in the initial state of the database. If only AEI method is used to adjust, it will take a long time, which is not conducive to the stability of service quality In order to speed up the optimization and adjustment of ciphertext segments, other methods such as equi width, equidepth or QOB can be used to divide several ciphertext segments during initialization, and then AEI method can be used for dynamic adjustment and optimization with the loading of data and services.

2.3 Implementation of Data Query Under the Guidance of Fuzzy Theory

Fuzzy sensor network is formed by replacing traditional sensors with fuzzy sensors in traditional sensor networks. Its network composition is the same as that of traditional wireless sensor networks. The data retrieval system of fuzzy sensor network is shown in Fig. 4.

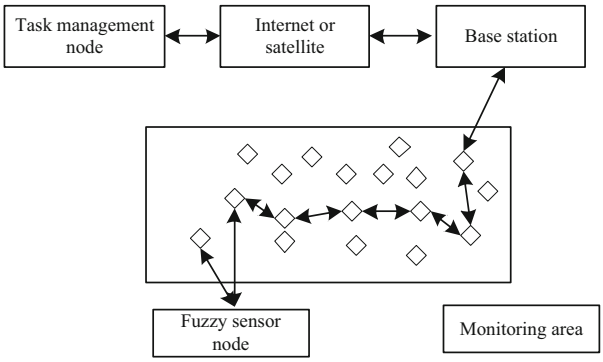

Fig. 4. Fuzzy sensor network data retrieval system

Sensor network is composed of a large number of micro sensor nodes deployed in the monitoring area. The nodes form a multi hop self-organizing network system through wireless communication. The network nodes can cooperatively sense, collect, process and transmit the information of the perceived target in the network coverage area, and send it to the information collector. Fuzzy sensor network can enable people to obtain a large amount of detailed and reliable information at any time, place and under any environmental conditions, so as to greatly expand the functions of the existing network. Fuzzy sensor network usually refers to a wireless network composed of a group of nodes with embedded processors, sensors and wireless transceiver, which collects and processes the target information in the network coverage area through the cooperative work of nodes. The sensor node is deployed in a target area. The information measured by the sensor node, such as temperature, humidity, light, pressure and speed, is transmitted to the convergence point through multi hop, connected through the convergence point, and finally connected to the task management node. The task management node has a man-machine interface, which can intervene, remote control and manage. Convergence point is a system with strong communication and computing power. In some documents and projects, it is also called base station.

Like the general sensor, the fuzzy sensor can sense the measured value determined by the sensing element, but the fundamental difference is that it can output not only the numerical value, but also the linguistic symbol. Therefore, the fuzzy sensor must have a numerical symbol converter. The fuzzy sensor should be able to exchange information with the superior system, so the communication function is the basic function of the fuzzy sensor. The basic function of fuzzy sensor determines its basic structure and implementation method. The basic structure of fuzzy sensor data management is shown in Fig. 5.

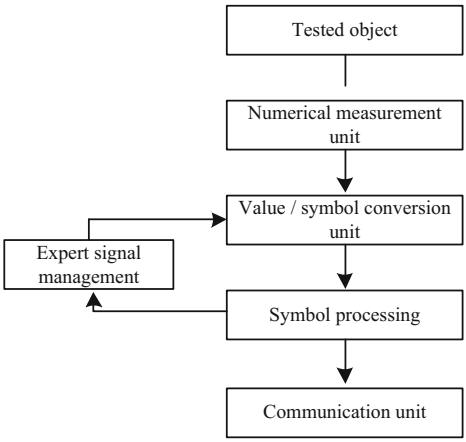

Fig. 5. Basic structure of fuzzy sensor data management

In order to effectively use the dynamic sensing data and accurately express the unique query of sensor networks, it is necessary to study and design a query processing system that can reflect the characteristics of sensor networks, so as to lay a foundation for query optimization and query processing. In the discussion of this paper, the specific uncertainty model is not important. What users need is the possible value of the known data in the query interval. This paper is interested in the query of some dynamic attributes L and the set S of objects in the database. Although the model in this paper can be extended to other domains, this paper still assumes that L is the attribute of a perceived object. For example, very simple integers and coordinates. In this paper, Si is used to represent the ith fuzzy sensor in S (Fig. 6).

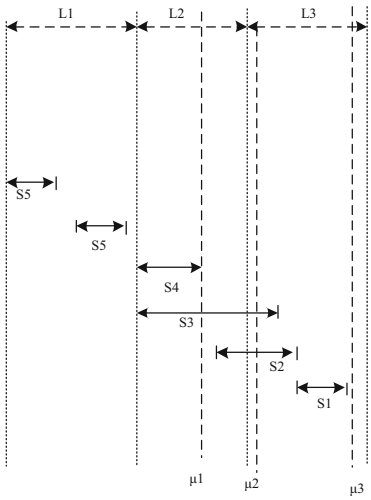

Fig. 6. Example of fuzzy data sensing query

In order to fully describe the query classification method proposed in this paper, a specific query example is given. For a given initial network, the network distribution of the encoded partition center node is obtained. For each partition, a point closest to the diagonal is selected as the storage data node, and the data in the partition will be transmitted to the node. The update mechanism adopts FIU update algorithm, that is, when the data value of fuzzy sensor changes beyond a certain threshold, update the value of database. After the slicing is completed, according to the method in Chapter 4, the virtual topology of the sliced network is generated, that is, K paths with the lowest communication cost are found from the base station to each central node and any pair of chip central nodes, and the network virtual topology is formed. The data between any nodes in the virtual topology will be saved in the routing table of the base station.

To sum up, the design of data query algorithm based on fuzzy theory is completed. The algorithm considers that the membership interval of the data generated by the fuzzy sensor is uncertain. Therefore, it divides a plurality of query areas, and returns the data characteristics of each area to the base station according to the query route. The base station obtains the final fuzzy theory value through sorting, so as to achieve the data query goal. The specific query process is shown in Fig. 7.

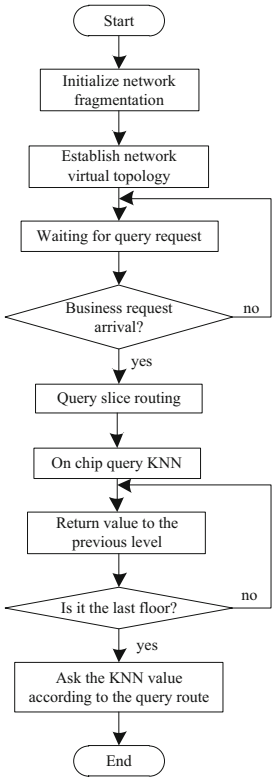

Fig. 7. Flow chart of data query algorithm based on Fuzzy Theory

3 Experiment and Result Analysis

In order to verify the effectiveness of the data query method based on fuzzy theory, the following simulation experiments are designed.

In the experiment, a prototype system is developed using C + technology to realize the development of data query methods. The main development tools are C + + 6.0 and SQL server. 54 sensors are set in the local area network to collect network data. The background database is built using sqlserver2000 technology. The operating environment is: WindowsXP, Pentium (R) d28gcpu, 1g, 160g hard disk. Planetsim is selected as the experimental simulation tool. The node size in the wireless network is set to change within the interval [0, 500], and 5000 query requests are generated randomly for each experiment.

In order to avoid the singleness of the experimental results, this method is compared with the traditional data optimization query method based on ant colony optimization and the data query method based on compression. The test comparison results when the data scale is 150000, 300000, 450000 and 600000 are as follows. In each figure, PQ represents the data optimization query method based on ant colony optimization, RQ represents the data query method based on compression, and FKNN represents the method in this paper.

The query response time results are shown in Fig. 8 and Fig. 9.

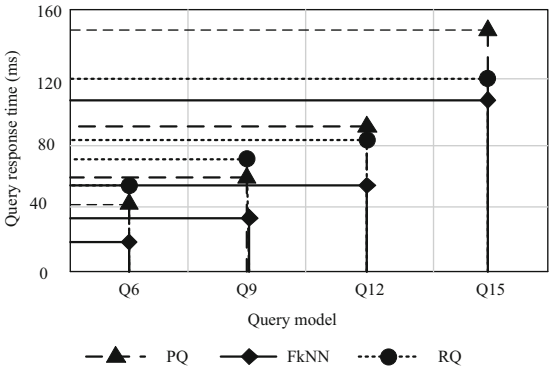

Fig. 8. Data query response time of each region

Because the network of this experiment is a small network composed of 54 sensors, the following methods are used to test the recall and precision of the proposed method: for each query instance, a corresponding data set is generated, including 24 appropriate sensor tuple combinations related to the query, These sensor tuple combinations are formed by removing duplicate tuples and appropriately adding randomly selected sensor tuples from the first 8 relevant query results obtained by the three query methods mentioned above. Then, the user selects 8 sensor tuples most related to the initial query, and tests the recall and precision of different methods on this basis. The results are shown in Fig. 10 and Fig. 11.

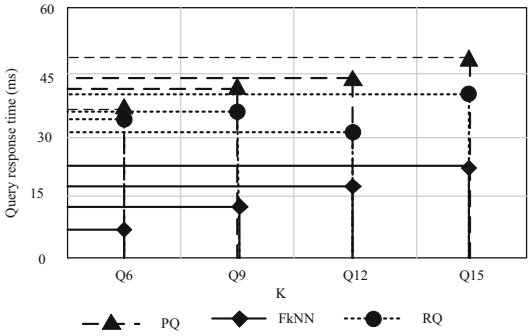

Fig. 9. Query response time when feature value changes

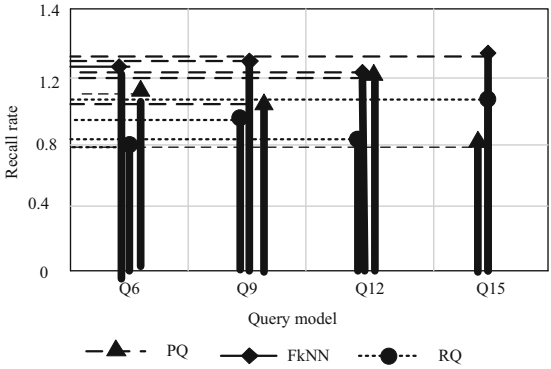

Fig. 10. Data query recall test results

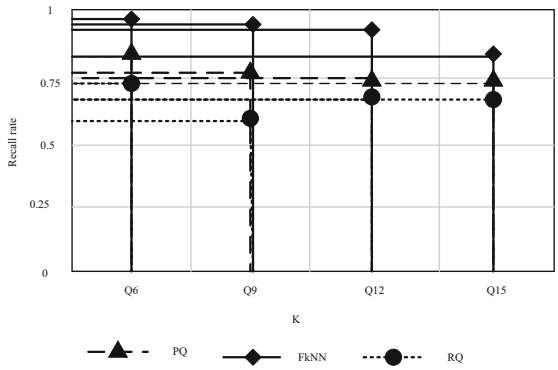

Fig. 11. Recall rate of changing data under interference environment

Based on the above analysis of the experimental results, it can be seen that the fuzzy theory query method proposed in this paper is better than the two traditional query methods in query response time and recall, which also shows that the query method proposed in this paper is feasible.

4 Conclusion

With the development of intelligent database, the processing of fuzzy data and its application in computer will have a better promotion and development space. Therefore, this study proposes a data query optimization method based on fuzzy theory.

In this study, the features of the data to be queried are identified by fuzzy theory. After the features are classified, multiple query areas are divided, and the data features of each area are returned to the base station according to the query route. The base station obtains the final fuzzy theory value through sorting, so as to achieve the data query goal. Through experimental verification, it can be seen that after the application of this method, the data query time is less, and the application effect in recall is also better than the traditional method.

References

1. Xu, B.: Massive incomplete data approximate query system based on improved K-nearest neighbor algorithm. Mod. Electron. Tech. **44**(15), 177–181 (2021)
2. Teng, Z., Liao, Z.: Differentiated service mechanism for data query on named data networking. Comput. Eng. Appl. **55**(9), 17–25+86 (2019)
3. Gao, J., Yang, F.: Semi-structured data query optimization algorithm based on swarm intelligence. Comput. Simul. **38**(8), 381–385 (2021)
4. Lian, J., Fang, S., Zhou, Y.: Model predictive control of the fuel cell cathode system based on state quantity estimation. Comput. Simul. **37**(07), 119–122 (2020)
5. Su, J., Xu, R., Yu, S., Wang, B., Wang, J.: Idle slots skipped mechanism based tag identification algorithm with enhanced collision detection. KSII Trans. Internet Inf. Syst. **14**(5), 2294–2309 (2020)
6. Su, J., Xu, R., Yu, S., et al.: Redundant rule detection for software-defined networking. KSII Trans. Internet Inf. Syst. **14**(6), 2735–2751 (2020)
7. Ma, Z., Yuan, H., Gu, Y., et al.: Research and implementation of document-relational data query execution technology. J. Front. Comput. Sci. Technol. **14**(08), 1315–1326 (2020)
8. Yun, W.: Improved Simulation of Large-Scale Hybrid Network Database Fuzzy Query Algorithm. Computer Simulation **37**(05), 246–249 (2020)
9. Lu, S., Chen, H.: A survey on data query optimization with machine learning. Wirel. Commun. Technol. **29**(04), 5–10 (2020)
10. Zhao, Y., Hu, L.: Design of preference query system based on linked data in open environment. Comput. Technol. Dev. **30**(09), 7–11 (2020)