

A Knowledge-based Method of Multilevel Query in Relational Database

Xu Jingke, Sun Huanliang, Wang Yonghui, Song Xiaoyu, Wang Shoujin

School of Information and Control Engineering
Shenyang Jianzhu University,
Shenyang, Liaoning, 110168, China

Abstract—There is a kind of query that users tend to use fuzzy concepts to express the query requirements. In the relational database, the query based on concepts as inputs cannot be directly executed. The basic method converts users' fuzzy concepts into precise concepts based on the domain knowledge. In practice, a fuzzy condition is often composed of the high and multilevel concepts, so its conversion of the concepts is very complex, and its efficiency is poor. A new knowledge-based method of multilevel query is proposed in this paper, and it is achieved on generalized data. Experiments show that the query efficiency of the method proposed in the paper is better than the basic method on multiple levels. It may provide efficient methods for the query about housing products in rural town, and it can meet the needs of customers with different knowledge, and it has higher practical value.

Keywords—Knowledge Base, Multilevel Query, Generalized Rule, Relational Database

I. INTRODUCTION

In the e-commerce system users do not know performance of commodities described by numeric values, but they can provide conceptual descriptions. The query based on concepts as inputs is named conceptual query. [1]. If a user needs to know "which product is the higher quality?", then the traditional database query cannot deal with this problem, because the "higher" is a fuzzy concept. This kind of query can be achieved by the knowledge-based method of data query [2, 3]. The basic method of conceptual query converts user's fuzzy concept into standard SQL statement based on the domain knowledge.

In conceptual query, the concepts are often hierarchical. It may contain arithmetic rules between levels [4, 5]. For example, a user needs to know "which product has the higher cost performance?", the "cost performance" is higher than "quality" and "price", its concept is calculated with the arithmetic rules on the fuzzy concept of the "quality" and "price". If the query still uses the basic method to convert the concepts into the SQL statement, then it is very complex, because the condition is composed of the higher and multilevel concepts. When using the knowledge to convert the concept into the SQL statement, it needs to calculate the higher concept with the lower concepts of several attributes according to certain rules. So it has multiple connections with itself, the efficiency of the multilevel query is poor, and the more complex the relationship is, the poorer its efficiency is.

In order to improve the efficiency of the multilevel query, a knowledge-based query method on multilevel generalized data is proposed.

II. RELATED WORKS

A concept hierarchy defines a sequence of mappings from a set of low level concepts to higher level, more general concepts [6]. The concept hierarchy in the paper belongs to the rule-based hierarchy, and the mappings need to be calculated according to certain rules. Concept hierarchy is important means of expressing domain knowledge. It can effectively transfer domain knowledge, and improve interesting and effectiveness of query results.

The intelligent query system must have the appropriate fuzzy knowledge base [7] to support. The knowledge base is the premise and basis for the conceptual query. By establishing a knowledge base of experts which the framework of the conceptual query is based on, it is one way to achieve conceptual query. In recent years, more and more research on database query is devoted to generalized process of the fuzzy conditions. This methods based on knowledge can translate the fuzzy query into the precise query, which didn't involve the query condition on multilevel fuzzy concepts, this paper mainly studies the multilevel conceptual data query problem.

III. PROBLEM DEFINITIONS

Definition 1 Knowledge base K : Let a relation $K(Kid, Value, Gid, Low, Up)$ express the knowledge, Kid is the primary key of the knowledge, $Value$ is the value of the knowledge, $Gid \in P(Pid, Gid, Value)$ is the field bound by the knowledge, Low is the lower limit, and Up is the upper limit. A tuple in K can describe a piece of knowledge.

Definition 2 Conceptual relation PE : Given a relation P and a knowledge base K , a generalized rule can be described by a piece of knowledge $k \in K$, According to the generalized rule k everyone in P is processed, and a conceptual relation is generated, which is expressed as $PE(Pid, Gid, Value)$. If p_e is a tuple of the relation PE , then it satisfies formula 1.

$$\begin{cases} p_e.Gid = k.Gid \\ p_e.Value = k.Value \mid (T.Value \geq k.Low) \wedge (T.Value \leq k.Up) \end{cases} \quad (1)$$

Here, $T.Value$ may be computed by predefined rules.

Definition 3 Knowledge-based multilevel query: On a table p_e whose relational schema is $PE (Pid, Gid, Value)$ a query Q is given in the form of $Q=C_1 \wedge \dots \wedge C_k$, here each $C_i (i=1, \dots, k, k \leq n)$ is a basic query condition. If there is at least one basic query condition contained in the conceptual relation PE , then the query Q is called knowledge-based conceptual query, and if a condition is on the concept at several levels, then it is called knowledge-based multilevel query.

IV. THE KNOWLEDGE-BASED MULTILEVEL QUERY

Taking data query about housing products in rural town for example (Table 1), a concept hierarchy of the product's attributes, the knowledge base, and the knowledge-based generalized data are given.

TABLE I PRODUCT FUNCTIONS

ID	D_{11}	D_{12}	D_{21}	D_{22}
P01	1.9	2.1	3.1	2.1
P02	2.4	2.4	3.4	1.4
P03	1.7	2.1	3.1	3.1
P04	3.4	2.4	3.4	3.4

A. The Knowledge-based Generalized Rule

A knowledge-based generalized rule consists of a concept hierarchy and a knowledge base, by which the generalized data of initial data is generated.

As shown in Figure. 1, it is the concept hierarchy of the product's attributes. Here identification codes of the attributes reflect their hierarchical relationship, the concept D_i which has identification code "01" is a higher level concept, and its lower level concept is the concept D_{11} which has identification code "0101", denoted as $D_{11} < D_i$; the mapping rule from a set of low level concepts to higher level is implemented by formula 2, here the value of $D_{ij} \cdot w$ is the weight of the attribute D_{ij} .

$$D_i \cdot Value = \sum_{D_{ij} < D_i} D_{ij} \cdot Value * D_{ij} \cdot w \quad (2)$$

In order to achieve the conceptual query, the domain knowledge base needs to be established, and a knowledge base K (as shown in table 2) used in the query is given according with the definition 1. Here, $Value$ is the value of the knowledge, expresses the fuzzy concepts such as level "A", "B", "C" "D", and is denoted as number 95, 85, 75 and 65 for the multilevel process. The knowledge base which exists independently can be updated constantly, ensuring reliability and feasibility of the knowledge [8].

TABLE II KNOWLEDGE BASE K

Kid	Value	Gid	Low	Up
K01	95	0101	1.0	1.5
K01	85	0101	1.5	2.0
K01	75	0101	2.0	2.5
K01	65	0101	2.5	5.0
K02	95	01	90	100
...

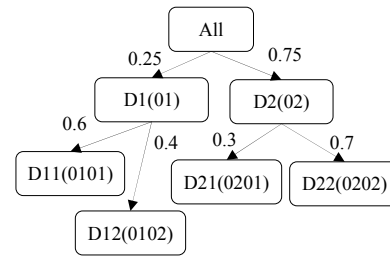


Figure 1. concept hierarchy

B. The Knowledge-based Generalized Data

In order to make the query's versatility and scalability better, product performance data shown in table 1 can be used to represent as initial data P according with define 1. Here, Pid is the primary key, Gid is the identification code of the attribute (the identification code "0101" is used to express the attribute D_{11}), $Value$ is the value of the attribute (such as D_{11}). One tuple in P includes an attribute of product and its value, and some tuples can describe one product.

In knowledge base K , the knowledge encoded as "K01" can be used to treat tuples whose identification code is "0101" in P on the lower concepts, and other tuples can be of similar treatment. The generalized data on low level of relation P is PE as shown in table 3.

TABLE III GENERALIZED DATA ON LOW LEVEL PE

Pid	Gid	Value
P01	0101	85
P01	0102	75
P01	0201	75
P01	0202	85
P02	0101	75
...

At the same time, because attributes is hierarchical, so knowledge formed by relationship between attributes is also hierarchical, as well as generalized data. Firstly, the weights of lower concepts contained on higher level should be obtained (Figure. 1), and then the weighted sum of generalized data PE on low level is computed according to formula 2. Finally, after PE is treated by the knowledge such as "K02" in K , the generalized data $PE2$ on higher level is obtained (table 4).

TABLE IV GENERALIZED DATA ON HIGH LEVEL PE2

Pid	Gid	Value
P01	01	85
P01	02	85
P02	01	75
...

C. The Knowledge-based Multilevel Query on

Generalized Data

In the knowledge-based conceptual query, If the query condition is only on the concept at a level, then the query is called single-level query; if the condition is on the concept at several levels, then it is called multilevel query. Here, the query about housing products in rural town is taken as an example to show how to use the method in the paper for achieving multilevel conceptual query.

Example: The User wants to query for products whose attribute D_1 is "B" and D_{21} is "C".

The query

$$\sigma_{Gid='01' \text{ And Value}=85}(P) \cap \sigma_{Gid='0201' \text{ And Value}=75}(P)$$

on the basic relation is denoted as Q . If the query is directly executed on the relation P , then the count of records returned is zero, and the eligible result set cannot be visually found. Therefore, it needs to convert the condition into standard SQL statement (as follows) according to the knowledge base K .

```
Select a.Pid
From
(Select Pid,K.Value From P,K Where P.Gid = '0101' And P.
Gid =K. Gid And P.Value>=K.Low And P.Value<K.Up) as a ,
(Select Pid,K.Value From P,K Where P.Gid = '0102' And P.
Gid =K. Gid And P.Value>=K.Low And P.Value<K.Up) as b ,
(Select Pid,K.Value From P,K Where P.Gid = '0201' And P.
Gid =K. Gid And P.Value>=K.Low And P.Value<K.Up) as c
Where
(a.value*0.6+b.value *0.4>=80) And (a.value *0.6+b.value
*0.4<90) And (c.value=75)
```

And a. Pid =B. Pid And b. Pid =c. Pid

Group By a. Pid

The query

$$\sigma_{Gid='01' \text{ And Value}=85}(PE2) \cap \sigma_{Gid='0201' \text{ And Value}=75}(PE)$$

on the conceptual relation is denoted as QE , and the relevant SQL statement is as follows.

```
Select Pid
From PE2
Where Gid='01' And Value=85 And Pid In(Select Pid From
PE Where Gid='0201' And Value=75)
Group By Pid
```

From the example presented above, it may be seen that the basic method to achieve the conceptual query on multilevel concepts needs to convert the conditions into the SQL statement, and its efficiency is poor. But the proposed method has the simple and efficient advantage under certain conditions, and the more complex the concept hierarchy is, the more obvious the advantages are.

V. EXPERIMENTS

In this paper there are 7 synthetic databases used for the experiment. They have the schema according with definition 1-2, whose indexes are established on the primary key. The random housing products are separately stored in each database, whose count is from 10000 to 70000. The initial data in P is generated into the generalized data "PE" and "PE2" according to the knowledge K (table 2) and the

concept hierarchy (Figure. 1). The experiments aim at comparing the efficiency with the queries.

In table 5 the CPU time is compared between Q and QE in the data sets which are different sizes, and the CPU time of QE is always smaller than Q with the same selection scale. It shows that the query efficiency of the method proposed in the paper is better than the basic method in the conceptual query on multilevel concepts. Its efficiency is improved by at least 40 percent compared with the basic method in this experimental environment. It increases the cost of data storage, and the storage efficiency is affected by the multilevel generalized data. In the application of giving priority to query efficiency, the knowledge-based multilevel query has certain advantages.

TABLEV COMPARISON OF QUERY EFFICIENCY BETWEEN Q AND QE

Data sets (10000)	1	2	3	4	5	6	7
Q CPU time(ms)	483	400	326	423	590	660	816
QE CPU time (ms)	26	100	183	230	316	370	340
Selection Scale (%)	1.1	1.1	1.1	1.1	1.1	1.1	1.1
Efficiency Increase (%)	95	75	44	46	46	44	58

VI. CONCLUSIONS

A knowledge-based query method on multilevel generalized data is proposed in the paper. Taking housing products in rural town for example, the generalized data of initial data is generated by knowledge base, and then the query is achieved on generalized data. Experiments on the synthetic data sets show that the query efficiency of the method proposed in the paper is obviously better than the basic method on multiple levels. The knowledge-based multilevel query will provide efficient methods for the query about housing products in rural town, and it is able to meet the needs of customers with different knowledge.

ACKNOWLEDGEMENTS

The work is supported by the National Technology R&D Program in the 11th Five year Plan of China (2008BAJ08B08), the National Natural Science Foundation of China (61070024).

REFERENCES

- [1] TANG Lin-yan. Data Warehouse Query and Implementation of Intelligent Query Programs [J]. Computer Engineering and Applications, 2000, (8):114-116.
- [2] CHEN Yi-fei, YE Xiao-ling, ZHANG Ying-chao. Fuzzy Queries Based on EPTV in Relational Databases [J]. Computer Engineering, 2010, 36(4):24-27.
- [3] ZHANG Ying-chao, CHEN Yi-fei, YE Xiao-ling, et al. Weighted fuzzy queries in relational database[C] //Lecture Notes in Computer Science. Berlin: Springer, 2005: 430-440.
- [4] ULRICH B, JOSEFK, SUSANNE S. Flexible query answering using distance-based fuzzy relations[C] //Lecture Notes in Computer Science. Berlin: Springer, 2006: 207-228.
- [5] MA Zong-min, YAN Li. Generalization of strategies for fuzzy query translation in classical relational databases[J]. Information and Software Technology, 2007, 49(2): 172-180.

- [6] CUI Xin-chun, HAN Li-li. An Efficient Algorithm of Mining Multi-Level Fuzzy Association Rules [J]. Computer Engineering and Applications, 2004, (10):187-189.
- [7] CAI Xiang, XU Dong-ping, ZHAO Yu-wei. Research of general data mining system based on knowledge intelligent query [J]. Computer Engineering and Design, 2006, 27(8): 1382-1384.
- [8] YUAN Lu, LIU Feng, LIAO Jie. Land-use data integration based on knowledge[J]. Computer Applications, 2010, 30(12):3385-3387.