Future Generation Computer Systems 82 (2018) 698–706

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Mining and updating association rules based on fuzzy concept lattice

Caifeng Zou [a,b], Huifang Deng [a], Jiafu Wan [c,*], Zhongren Wang [d], Pan Deng [e,f]

[a] School of Computer Science and Engineering, South China University of Technology, Guangzhou, PR China
[b] College of Computer and Design, Guangdong Mechanical and Electrical College, Guangzhou, PR China
[c] School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou, PR China
[d] School of Mechanical and Automotive Engineering, Hubei University of Arts and Science, Xiangyang, PR China
[e] Center for Data Science, Peking University, Beijing, PR China
[f] Beijing Institute of Big Data Research, Beijing, PR China

HIGHLIGHTS

- Proposes an algorithm for mining and updating association rules based on fuzzy concept lattice.
- Extends the existing incremental construction algorithm of precise concept lattices.
- Uses the pruning technology to improve the construction algorithm.
- Adds the steps for generating and updating association rules.
- Reduces the computational workload.

ARTICLE INFO

Article history:
Received 13 June 2017
Received in revised form 26 October 2017
Accepted 11 November 2017
Available online 20 November 2017

Keywords:
Association rule
Fuzzy concept lattice
Construction algorithm
Data mining

ABSTRACT

An algorithm for mining and updating association rules based on fuzzy concept lattice is proposed. When a new attribute is added into the fuzzy concept lattice, it is not necessary to calculate all the frequent nodes and association rules. According to the incremental construction algorithm of fuzzy concept lattice, it is only necessary to deal with the new nodes that have changed. Therefore, the amount of calculation is reduced. The existing incremental construction algorithm of precise concept lattices based on attributes is extended so that it can be applied to fuzzy concept lattices. The pruning technology is used to improve the construction algorithm. The steps for generating and updating association rules are added. According to the extended algorithm, the fuzzy concept lattice can be constructed, and the corresponding association rules can be generated and updated at the same time. The experimental results show that the algorithm proposed in this paper greatly reduces the computational workload, and shortens the running time.

1. Introduction

With the continuous popularization of Internet and the rapid development of IoT (Internet of Things), big data attracts more and more attention [1]. Internet data have increased dramatically, which almost doubled every two years [2]. Big data includes not only information posted on the Internet, but also huge amounts of information generated by digital sensors on industrial equipment, cars, and meters [3–10].

Data Mining uses techniques and methods such as artificial intelligence, machine learning, statistics and database systems to explore valuable rules and information from large data sets [11,12]. Association rule extraction is an important research content in data mining field. Its purpose is to extract valuable association rules from massive data and display them intuitively.

Formal Concept Analysis (FCA) is a kind of data analysis tool different from traditional knowledge representation. Concept lattice is the key data structure in formal concept analysis, which can be used to represent classification and hierarchical relations between attributes [13]. It is an effective tool for data analysis and association rules extraction.

In the objective world, the information is mostly vague and indefinite, so the research on fuzzy concept lattice has important practical significance. With the change of objective objects, fuzzy concept lattices need to be adjusted constantly, and the association rules based on fuzzy concept lattices should be updated accordingly. This process of updating and modifying has certain rules. It is not necessary to recalculate all frequent nodes and association rules when the fuzzy formal context is partially changed. Only

* Corresponding author.
E-mail addresses: caifengzou@gmail.com (C. Zou), hdeng2008@gmail.com (H. Deng), mejwan@scut.edu.cn (J. Wan), xfu_wangzhongren@126.com (Z. Wang), 13811223436@139.com (P. Deng).

https://doi.org/10.1016/j.future.2017.11.018
0167-739X/

new nodes that have changed need to be processed. Thus the computational complexity is reduced.

This paper proposes an association rule mining and updating algorithm based on fuzzy concept lattice, which can not only construct fuzzy concept lattice, but also generate and update the corresponding association rules synchronously. The extraction of association rules based on fuzzy concept lattice enlarges the range of data processing, and is more suitable for mining uncertain data in reality. When the new attribute data is added into the fuzzy concept lattice, it is unnecessary to recalculate all frequent nodes and association rules. According to the incremental construction algorithm of fuzzy concept lattice, it is only necessary to deal with the new nodes that have changed. The incremental construction method of fuzzy concept lattices based on attributes is extended, and the related steps of generating and updating association rules are added. The algorithm proposed in this paper greatly reduces the computational workload.

2. Related work

Many researchers have made a thorough study of the theory and application of precise concept lattice in recent years. The extraction algorithm of association rules based on precise concept lattice has been widely used. However, the research on mining association rules based on fuzzy concept lattice is relatively few. Fuzzy concept lattice is an excellent tool for extracting association rules. Its Hasse diagram can display the partial order relationship in the fuzzy concept lattice succinctly and concisely, so it is conducive to mining the association rules quickly and intuitively.

2.1. Fuzzy set

Many objective phenomena are vague and have no definite boundaries. In the information society, the data collected in many real systems are inaccurate. If these uncertain information cannot be properly analyzed, it will probably lead to a great error between the inference result and the objective fact. Many researchers have been trying to find a reasonable way to deal with uncertain data. In 1965, Zadeh proposed the fuzzy set as an extension of the classical set theory [14]. According to the definition in mathematics, the elements in fuzzy sets have membership value. Salii defined a more general structure in the abstract algebra context, which was called L-relation [15]. Fuzzy relations are now widely used in different fields, such as linguistics, decision systems, and cluster analysis. The fuzzy set theory uses a membership function to describe the fuzzy relation. The range of the membership function value is the unit interval [0, 1].

Fuzzy sets can be represented as (U, m), wherein the domain U is a set, and $m : U \to [0, 1]$ is a membership function. For each $x \in U$, the value of $m(x)$ is called the membership degree of x in (U, m). For the finite set $U = \{x_1, \ldots, x_n\}$, the fuzzy set (U, m) is usually represented by $\{m(x_1)/x_1, \ldots, m(x_n)/x_n\}$.

For a given domain U, a mapping $m_A : U \to [0, 1]$ can determine a fuzzy subset A, whose representation is as follows:

$$A = \begin{cases} \displaystyle\sum_{x_i \in U} \frac{m_A(x_i)}{x_i}, & U \text{ is a finite set } \{x_1, \ldots, x_n\} \\ \displaystyle\int_{x \in U} \frac{m_A(x)}{x}, & U \text{ is an infinite set} \end{cases} \quad (1)$$

Wherein m_A is the membership function, and $m_A(x)(x \in U)$ indicates the membership degree of x to A. Note that $\sum_{x_i \in U} \frac{m_A(x_i)}{x_i}$ is not an expression of the sum of fractions, but merely a sign.

2.2. Fuzzy concept lattice

There are many vague concepts in people's thinking, such as young, smart, warm, pretty, etc. In the real world, there are a lot of fuzzy phenomena [16]. Fuzzy theory and formal concept analysis can be combined to deal with uncertain information [17]. In 1994, A. Burusco and R. Fuentes proposed the fuzzy concept lattice theory [18], which is the combination of concept lattice and Zadeh fuzzy set theory.

Liu et al. [19] and Qiang et al. [20] discussed the construction algorithms and applications of fuzzy concept lattice. Jaoua and Elloumi [21] took advantage of implication operator to generate the fuzzy concept lattice, studied the concept lattice with varying accuracy, and introduced the variable threshold concept lattice. Tho et al. [22] studied the combination of the fuzzy set theory and FCA, and discussed the technique of Fuzzy Formal Concept Analysis (FFCA). This technique can be used to calculate the similarity between the fuzzy formal concepts. Sun and Qin [23], Liu et al. [24] used residual implication and different implication operators to generate fuzzy concept lattice. Sun et al. [25] discussed an fuzzy concept generation method based on alternate objects and attributes. Chen et al. [26], Chen and Luo [27] studied the generation approach of interval-valued fuzzy concept lattice.

2.3. Attribute-based incremental construction algorithm of precise concept lattices

The basic idea of the incremental construction algorithm based on attributes is to initialize the concept lattice with a top node firstly. Then the attributes are inserted into the concept lattice one by one. When each attribute is inserted, all concept nodes in the concept lattice should be traversed. The intersection between the set of objects of the attribute to be inserted and the extensions of all nodes in the concept lattice should be calculated. Further processing based on the calculation results should be done. Generally, a bottom-up approach is adopted to traverse the concept lattice in descending order of the cardinalities of intensions [28].

Suppose that $C(A, B)$ is a node on a concept lattice, and the new attribute to be inserted and its object set are $(g(d), d)$. The extensions need to be compared when the attribute is inserted. According to the comparison results, the following two situations should be dealt with accordingly [29]:

(1) *Updated nodes*
If $A \cap g(d) = A$, then C is called the updated node, and C is updated to $(A, B \cup d)$.

(2) *New added nodes*
If $A \cap g(d) \neq A$, and $\neg\exists C_S(A_S, B_S)$ satisfies that $C_S(A_S, B_S)$ is a child node of $C(A, B)$ and $A_S \supseteq A \cap g(d)$, then the new added node $(A \cap g(d), B \cup d)$ is inserted.

Both the updated nodes and the new added nodes are called the newborn nodes. For the node $C(A, B)$, if $\exists N$ satisfies that N is a newborn node and $C > N$, and $\neg\exists H$ satisfies that H is a newborn node and $C > H > N$, then N is the closest child node of C, which is denoted by S^*. When it is determined whether a new added node is needed, the range of nodes that need to be traversed can be reduced from all the child nodes of C to the closest child nodes [29]. The above second situation can be improved as follows: If $A \cap g(d) \neq A$, and $\neg\exists S^*(A_{S^*}, B_{S^*})$ satisfies that S^* is a closest child node and $A_{S^*} \supseteq A \cap g(d)$, then the new added node $(A \cap g(d), B \cup d)$ is inserted.

The method of calculating the child nodes of the new added node $(A \cap g(d), B \cup d)$ is as follows: If $C(A, B)$ has a closest child node S^*, then S^* is a child node of $(A \cap g(d), B \cup d)$. In addition, the unique parent node of the new added node $(A \cap g(d), B \cup d)$ is $C(A, B)$. According to the above properties, H. Zhi proposed