Knowledge-Based Systems 54 (2013) 53–65

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

MOGA-based fuzzy data mining with taxonomy ☆

Chun-Hao Chen [a], Ji-Syuan He [a], Tzung-Pei Hong [b,c,*]

[a] Department of Computer Science and Information Engineering, Tamkang University, Taipei 251, Taiwan
[b] Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan
[c] Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung 804, Taiwan

ARTICLE INFO

Article history:
Available online 12 September 2013

Keywords:
Data mining
Fuzzy sets
Fuzzy rules
Multi-objective genetic algorithm
Taxonomy

ABSTRACT

Transactions in real-world applications usually consist of quantitative values. Some fuzzy data mining approaches have thus been proposed for deriving linguistic rules from such transactions. Since membership functions may have a critical influence on the final mining results, several genetic-fuzzy mining approaches have been proposed for mining appropriate membership functions and fuzzy association rules at the same time. Most of them, however, focus on a single level and consider only one objective function. This paper proposes a multi-objective multi-level genetic-fuzzy mining (MOMLGFM) algorithm for mining a set of non-dominated membership functions for mining multi-level fuzzy association rules. The algorithm first encodes the membership functions of each item class (category) into a chromosome according to the given taxonomy. Two objective functions are then considered. The first one is the knowledge amount mined out at different levels, and the second one is the suitability of membership functions. The fitness value of each individual is then evaluated using these two objective functions. After the evolutionary process terminates, various sets of membership functions can be used for deriving multi-level fuzzy association rules according to decision-makers. Experimental results on the simulated and real datasets show the effectiveness of the proposed algorithm.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Data mining is most commonly used to derive association rules from transaction data [1]. A variety of mining approaches based on the Apriori algorithm [1] have been proposed [4,32]. Most of them focus on finding association rules on a single level, and many approaches have been proposed for mining multi-level association rules to discover more specific and important knowledge from data [14,33].

Most of the approaches mentioned above focus on binary-valued transaction data, but data in real-world applications usually consist of quantitative values. Fuzzy data mining algorithms have thus been proposed for handling quantitative transactions and mining fuzzy association rules [5,17,26,35]. As for multi-level fuzzy mining, according to the given fuzzy generalization hierarchies, Lee proposed a generalized fuzzy quantitative association rule mining algorithm [23], and Hong et al. proposed a multi-level fuzzy association rule mining approach [18]. Their method first transforms quantitative transactions into fuzzy values using the predefined membership functions and taxonomy, and then a top-down progressively deepening approach is used to find large itemsets and rules. Additionally, in order to reduce the time complexity, only the linguistic term with the maximum cardinality of each item is used in the later mining processes. In [21], Kaya and Alhajj proposed a weighted fuzzy rule mining approach based on Hong et al.'s approach. In 2008, Lee et al. proposed a fuzzy mining algorithm for discovering generalized fuzzy association rules with multiple supports of items to extract implicit knowledge from quantitative transaction data [27].

These earlier approaches all assume that the membership functions are known in advance. However, the given membership functions may have a critical influence on the final mining results. In the past decade, some genetic systems and approaches for knowledge engineering have been proposed, such as genetic tuning, genetic KB learning, and genetic rule learning [12,24]. Various genetic-fuzzy mining (GFM) approaches have also been proposed to derive appropriate membership functions and mining fuzzy association rules [2,6,16,20,22]. Kaya and Alhajj proposed a genetic algorithm (GA)-based approach to derive a predefined number of membership functions to obtain the maximum profit within a user-specified interval of minimum supports [20]. Hong et al. also proposed a GFM algorithm for extracting both association rules

☆ This is a modified and expanded version of the paper "MOGA for Multi-Level Fuzzy Data Mining," presented at *The IEEE International Conference on Granular Computing*, 2012.

* Corresponding author at: Department of Computer Science and Information Engineering, National University of Kaohsiung, Kaohsiung 811, Taiwan. Tel.: +886 75919191.

E-mail addresses: chchen@mail.tku.edu.tw (C.-H. Chen), 600411861@s00.tku.edu.tw (J.-S. He), tphong@nuk.edu.tw (T.-P. Hong).

0950-7051/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.knosys.2013.09.002

and membership functions from quantitative transactions [16]. Alcala-Fdez et al. then modified this approach and proposed an enhanced approach based on the 2-tuples linguistic representation model [2]. Matthews et al. proposed a temporal fuzzy association rule mining with the 2-tuples linguistic representation [29]. In addition, multi-objective GFM approaches have been proposed [6,22,28].

However, these GFM algorithms focus on a single level. Few works have focused on multi-level GFM algorithms. In [7], the present authors proposed a multi-level GFM algorithm for mining membership functions and fuzzy association rules with taxonomy. Since decision-makers may consider different criteria in real applications, multi-objective optimization problems have become increasingly important. The present study thus proposes a multi-objective multi-level GFM (MOMLGFM) algorithm for mining sets of membership functions and fuzzy association rules. According to the given taxonomy, the algorithm first encodes the membership functions of each item class (category) into a chromosome. Two objective functions are then used to evaluate the fitness value of each individual. The first one is the summation of the large 1-itemsets of each item in all concept levels. The second one is the suitability of membership functions in each chromosome. After the evolution process terminates, sets of membership functions can be used for mining multi-level fuzzy association rules according to the different criteria of the decision-makers.

The remaining parts of this paper are organized as follows. The GA-based multi-objective optimization problem is stated in Section 2. The multi-objective genetic algorithm (MOGA)-based multi-level fuzzy data framework is described in Section 3. The components of the proposed approach are described in Section 4. The details of the proposed algorithm are given in Section 5. An illustrative example is presented in Section 6. Experiments to show the performance of the proposed algorithm are reported in Section 7. The conclusion and future work are stated in Section 8.

2. Ga-based multi-objective optimization problems

In traditional optimization problems, the goals to be achieved are usually transformed into fitness functions for maximization or minimization. Unfortunately, it is not easy to find the best fitness function for a problem in advance. In addition, several criteria may be considered in a real application, making multi-objective optimization problems increasingly important. Formally, a multi-objective optimization problem can be defined as follows:

$$\begin{aligned}&\text{Min/Max} \quad y = g(x) = (g_1(x), g_2(x), \ldots, g_m(x))\\&\text{subject to} \quad x = (x_1, x_2, \ldots, x_n) \in X \text{ and } y = (y_1, y_2, \ldots, y_m) \in Y\end{aligned}$$

where x is the decision vector, y is the objective vector, X represents the decision space, and Y represents the objective space. Several GA-based approaches have been proposed to solve such problems. For example, Schaffer proposed the vector-evaluated genetic algorithm (VEGA) to solve multi-objective optimization problems [31]. Fonseca et al. pointed out that VEGA has two problems [11]. The first one is that two non-dominated individuals are sampled at different rates. The second one is that the population tends to split into different species. They thus proposed a modified approach called MOGA that uses the extended rank-based fitness assignment [11] for solving the above two problems. They also defined three relationships among chromosomes, namely inferiority, superiority, and non-inferiority [11]. The MOGA strategy is used to find the set of non-inferiority solutions, also called Pareto-optimal solutions or the Pareto front. Fig. 1 explains the three relationships and Pareto-optimal solutions.

In Fig. 1, there are ten chromosomes and two objectives. The two objective values of a chromosome are represented by a data point in the figure. Take chromosomes C_1 and C_2 as an example. Chromosome C_2 is said to be inferior to C_1 since both the objective values of C_2 are worse than those of C_1. In this case, it is also said that C_2 is dominated by C_1. In other words, chromosome C_1 is superior to and dominates C_2. Chromosome C_1 is said to be non-inferior to C_7 and vice versa. In this case, both C_1 and C_7 are non-dominated points. The goal of MOGA is thus to find the non-dominated points (Pareto-optimal solutions). In this example, chromosomes C_1, C_7, C_8, C_9, and C_{10} are non-dominated points.

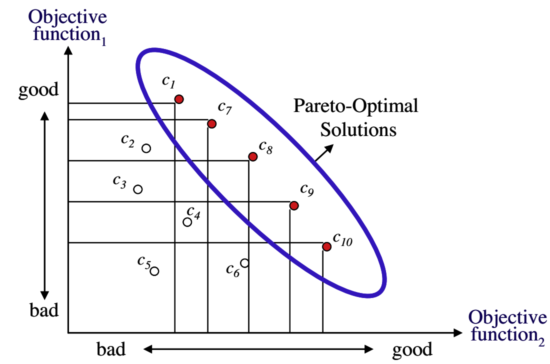

Fig. 1. Example of Pareto-optimal solutions.

3. Moga-based multi-level fuzzy data mining framework

This section proposes a MOGA-based multi-level fuzzy data mining framework for mining membership functions that are suitable for items with taxonomy. The derived Pareto-optimal solutions are then used to mine multi-level fuzzy association rules. The proposed framework is shown in Fig. 2.

The proposed framework consists of two phases, namely a multi-objective genetic-fuzzy membership function (MF) acquisition process and a multi-level fuzzy association rule mining process. The first phase generates and transforms the membership functions of each item category into a fixed-length string, which is known as a chromosome or an individual, according to the given taxonomy. The values of the objective functions of all chromosomes are then calculated for evaluating fitness values and executing genetic operations. The Pareto-optimal solution set is then used to keep the non-dominated solutions during the acquisition process. In the second phase, each Pareto solution (a set of membership functions) is used to mine multi-level fuzzy association rules. Note that various rule mining approaches can be used in this phase [18].

4. Components of proposed approach

This paper proposes the multi-objective multi-level GFM algorithm for deriving a set of membership functions for mining fuzzy rules. The details of the components in the proposed approach are described below.

4.1. Chromosome representation

It is important to encode membership functions as a string representation for GAs to be applied. Several possible encoding approaches are described in [2,8,25,30]. In order to effectively encode the associated membership functions, two parameters are used to represent each one, as in Parodi and Bonelli [30]. Thus, in multi-level fuzzy rule mining, membership functions applied to an item class (category) are assumed to be isosceles-triangle functions, as shown in Fig. 3.

Download English Version:

https://daneshyari.com/en/article/6862687

Download Persian Version:

https://daneshyari.com/article/6862687

Daneshyari.com