Applied Soft Computing 11 (2011) 4087–4096

Contents lists available at ScienceDirect

Applied Soft Computing

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

An approach for multi-objective categorization based on the game theory and Markov process

Wei-Yi Liu, Kun Yue*, Tian-Ying Wu, Mu-Jin Wei

Department of Computer Science and Engineering, School of Information Science and Engineering, Yunnan University, Kunming 650091, PR China

ARTICLE INFO

Article history:
Received 3 July 2010
Received in revised form 4 December 2010
Accepted 17 February 2011
Available online 9 March 2011

Keywords:
Multi-objective categorization
Shapley value
Coalitional games
Priority groups
Markov process

ABSTRACT

Realistic objects are not only described by their own attributes, but also described by their mutual relationships in a specific domain. By mainly considering the mutual associations among the given objects, in this paper we propose a method for multi-objective categorization based on the game theory and Markov process. We adopt Shapley value in coalitional games to measure the player's satisfaction degree in a group. We then give the concept of priority groups and an algorithm to combine small-size priority groups to large-size ones, and thus the efficiency of calculating the players' satisfaction degree can be improved. We further define a improving-replay Markov process to model the process of forming a reasonable payoff configuration. Accordingly, we give a simulation algorithm to obtain the desired payoff configuration to categorize players into groups by their satisfaction degrees. Finally, we give experimental results and performance studies to verify the efficiency and effectiveness of our methods.

1. Introduction

Realistic objects in a specific domain are not only described by their own attributes but also described by their mutual association established on business rules. According to people's intuition and general understanding, all relevant objects should be satisfied as possible with respect to different objectives when fulfilling a specific task on them. Discovering different groups, or called classes, is useful for pattern recognition, data preprocessing, association analysis, query optimization, etc. From the perspective of object behaviors under specific business rules, the ultimate groups should be determined by the associations among participating objects other than the attributes owned by the objects themselves. The associations among participating objects depend on their behaviors that generate groups. In this paper, with respect to the ultimate groups, we call the objectives that each object wants to achieve as multi-objectives or preferences (i.e., behavior tendencies), and we call the process of discovering different groups from given objects as multi-objective categorizations.

For example, three families $\{A, B, C\}$ will join touring parties to two tour objectives, *London* and *Paris*. The prices for the two tour objectives are given and each family has a preference for these two objectives respectively. Family A may intend to *Paris* together with family B and family C, but family B may intend to *London* together with family C. How do the three families join touring parties such that each family's preference is satisfied as much as possible under the given prices?

This example implies a general case of multi-objective categorization in real applications, in which the participating objects are associated by their behaviors for the certain task. Actually, there are many categorization problems as mentioned above in decision-making processes of an enterprise. The underlying methods for this kind of problems are desirable due to their wide occurrences and applications of knowledge discovery, data analysis, decision making, etc.

It should be noted that the multi-objective categorization mentioned above is different from the classical clustering in data mining and machine learning paradigms. It is well known that clustering is the process of grouping data into classes or clusters so that objectives within a cluster have high similarity in comparison to one another, but are very dissimilar to the objects in other clusters [8]. Various clustering methods have been proposed from various perspectives [11,26], in which the attributes of objects are considered as the basis for clustering. However, the associations among objects and the multiple objectives that each object should satisfy upon the ultimate groups have not been incorporated, although the demands for considering the associations as the multiple objectives are pervasive. Therefore, the classical clustering methods cannot be well suitable for the multi-objective categorization problems, and a novel approach is indispensable.

Game theory [22] is the study of mathematical models for conflict and cooperation between decision makers, and it has become an effective tool to describe strategic interactions in real world applications. Starting from the inherent characteristics of partic-

* Corresponding author. Tel.: +86 871 5033146; Fax: +86 871 5031598.
E-mail address: kyue@ynu.edu.cn (K. Yue).

1568-4946/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2011.02.023

ipating objects and ultimate groups, the influence of an object in a group is dependent not only on itself but also on the other ones. At the same time, conflicts among objects co-exist with the coordinations among them. This exactly forms a game [22] when we consider the multi-objective categorization. Particularly for the problems with game characteristics, Kleinberg et al. [12] advocate a utility oriented view of data mining driven by microeconomic considerations. In our study, inspirited by Kleinberg's work and based on the game theory, we will present an approach for multi-objective categorization. Fortunately, a coalitional game is a model of interacting decision-makers that focus on the behaviors of groups of participating players (i.e., objects), while each group of players is called a coalition [16,29]. Consequently for the multi-objective categorization, we can naturally model the desirable groups based on the inherence of coalitions and model the process of obtaining these groups based on the inherence of coalitional games. This is exactly our basic idea that will be interpreted in later discussions of this paper.

For various objectives, players have different combinations to compose several groups. In this paper, we call the intention of player i in a group as his satisfaction degree in the group. We notice that the strategies between player i and player j will generate conflicts, if player i intends to join a group g together with player j, but player j intends to join another group g'. From the above statements centered on obtaining the groups as coalitions, the following problems are quite deserving our attention:

- How to represent and compute a player's satisfaction degree in a group?
- How to solve conflicts between strategies of players?

For the first problem, we represent the combination of players' satisfaction degree in a group based on the idea of coalitional game. It is known that Shapley value [25] is an important value division scheme in coalitional game, and its main advantage is that the unique and fair solutions can be provided. Thus, in this paper we adopt Shapley value as a measure of the player's satisfaction, which is named satisfaction degree, in a coalition.

For the second problem, we assume that each player can get the chance to leave his current coalition and join another coalition for gaining the maximal payoff. The present coalitions of all players are looked upon as a payoff configuration in our paper. It is well known that Markov process [15] is a stochastic process whose future probabilities are determined by its most recent values. Fortunately, the future payoff configuration just depends on the present payoff configuration and is independent of the past configurations, which is consistent with the concept of Markov process. We will take the payoff configurations as Markov states and define the improving-replay Markov process to describe the transition process between payoff configurations and give an algorithm to obtain a desirable payoff configuration. In our work, the improving-replay Markov process describes the transition process of the payoff configurations. Each player in the obtained payoff configuration just belongs to one coalition and has the maximal satisfaction degree in the coalition. That is, the conflicts among players is solved and the players are categorized to coalitions according to their satisfaction degrees.

However, Shapley values cannot be computed in polynomial time [14]. Some approximation methods [29] have been developed pertinent to the problem of computational hardness of finding exact Shapley values. Because the number of combinatorial coalitions is very large, the computation of Shapley values for each player with respect to various coalitions is still a problem, even though some approximation methods for finding a Shapley value can be adopted. To improve the efficiency of calculating Shapley values, in this paper, we first give the definition of a priority group, in which the contribution of every player is larger than those of the players in other groups. Then we give an algorithm for combining small-size priority groups to large-size priority groups. Moreover, to guarantee the correctness of the algorithm, an order-preservation theorem is given at first.

As a way of solving conflicts, Nash equilibrium [21] is often used in a game activity. However, Nash equilibrium may not always be an optimal solution, and thus the conflicts cannot be well solved as desired. To find the desired solution to solving the players' conflicts, we propose an improving-replay Markov process, which is used to simulate the forming process of the desired payoff configuration. However, the Markov process is a stochastic process and the number of the Markov states is usually large. How to control the transition process of the Markov states is desirable to be considered. To control the transition process of the improving-replay Markov states, we define the concepts of the ration function and the absorption coefficient. Intuitively, most of the players should be willing to accept the payoff configuration that is used to categorize the players into appropriate coalitions according to their satisfaction degrees. Therefore, we give a simulation algorithm to find such a payoff configuration based on the improving-replay Markov process.

Generally speaking, the main contributions of this paper can be summarized as follows:

- We present the multi-objective categorization problem and the basic ideas by computing Shapley values.
- We give the algorithm for combining small-size priority groups to large-size priority groups, so that a player's satisfaction degree (i.e., Shapley values) in a group can be obtained efficiently.
- We define a improving-replay Markov process to model the process of forming a reasonable payoff configuration and give the approach for computing the max-absorbing states that form the reasonable payoff configuration, so that the conflicts between strategies of players can be solved.
- We implement our methods for obtaining the players' satisfaction degree in a group and categorizing players into groups according to their satisfaction degrees. Experimental results and performance studies verify the efficiency and effectiveness of our methods.

The remainder of this paper is organized as follows: Section 2 introduces related work. Section 3 gives basic ideas, basic concepts and problem statements. Section 4 gives the concept of priority group and the algorithm for computing their Shapley values. Section 5 gives the algorithm to obtain a reasonable payoff configuration based on the improving-replay Markov process. Section 6 shows experimental results and performance studies. Section 7 concludes and discusses our future work.

2. Related work

Clustering analysis is an important and classical aspect of data mining and machine learning [7–9,28]. In recent years, various clustering methods have been proposed from various perspectives [8,11,16], such as k-means, k-medoids, and DBScan. These methods were established mainly on the attributes owned by the participating objects themselves, instead of the associations (or called behaviors) among them.

Game theory involves cooperative and non-cooperative games [2,22,25]. The idea of cooperative games has been used in various aspects of the distributed artificial intelligent and multi-agent systems [24,27]. A Shapley value is a classical cooperative solution concept, and it provides a unique and fair solution to the coalitional game [25]. Unfortunately, Shapley values are difficult to compute