International Journal of Approximate Reasoning 81 (2017) 103–114

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Gini objective functions for three-way classifications ☆

Yan Zhang, JingTao Yao *

Department of Computer Science, University of Regina, Regina, SK S4S 0A2, Canada

ARTICLE INFO

Article history:
Received 30 June 2016
Received in revised form 21 October 2016
Accepted 7 November 2016
Available online 16 November 2016

Keywords:
Three-way classifications
Rough sets
Impurity
Gini coefficient
Objective function

ABSTRACT

The three-way classifications aim to divide the universe of objects into three disjoint regions, i.e., acceptance, rejection, and non-commitment regions. We can induce different types of classification rules from these regions. There exist different measures to evaluate the quality of regions. The partition of the three regions based on certain measures such as Gini coefficient is one of the challenges in three-way classifications. When using Gini coefficients to evaluate the impurities of three-way regions, there may exist contradiction on changing of various regions towards the preferred measure levels. The impurity of one region decreases at the expense of the increase of other regions' impurities when regions change. It is impossible to decrease one region's impurity without increasing the other regions' impurities. In this paper, we formulate Gini objective functions to balance the contradictions among the impurities of three-way regions. Three Gini objective functions, i.e., minimizing the overall impurity of three regions, minimizing impurities of immediate and non-commitment decision regions simultaneously, and minimizing impurities of acceptance, rejection and non-commitment regions simultaneously, are discussed in detail. These Gini objective functions express different preferred situations of three-way regions. The balanced three-way regions representing the trade-off among impurities can be obtained by finding the solutions to these Gini objective functions. An example shows how and what three-way regions are obtained by tuning impurities of these regions to satisfy certain Gini objective functions. It is suggested that with the proposed Gini objective functions more efficient and applicable three-way regions may be induced.

1. Introduction

Three-way decisions are formed based on the notions of acceptance, rejection and non-commitment [34]. Three-way classification is one of main formulations of three-way decisions [35]. Given U a finite nonempty set of objects and C a finite set of criteria, the three-way classifications divide U on C into three disjoint regions, i.e., positive, negative, and boundary regions [34]. In this paper, these three regions are called acceptance, rejection and non-commitment regions in order to distinguish three-way regions from rough set regions. The three-way classifications can be formulated from different models or theories, such as rough sets, fuzzy sets, shadow sets, and interval sets [5,34,36]. We focus on three-way classifications in the context of rough sets in this research.

☆ This paper is part of the virtual special issue on tri-partition, edited by Davide Ciucci and Yiyu Yao.

* Corresponding author.
E-mail addresses: zhang83y@cs.uregina.ca (Y. Zhang), jtyao@cs.uregina.ca (J. Yao).

http://dx.doi.org/10.1016/j.ijar.2016.11.005
0888-613X/

A critical research question of three-way classifications is to determine a better partition in dividing the universe U into three-way regions. In order to solve this problem, we need analyze some measures or criteria such as accuracy, coverage, cost, or uncertainty to evaluate these regions. These measures evaluate three-way regions from different views. The determination of regions based on the measures that are used to evaluate regions is one of the fundamental issues when applying three-way classifications [33]. When accuracy is emphasized in evaluating three-way regions, the three regions formulated by Pawlak rough set model are the best outcome as the highest accuracy of 100%. If the coverage of three-way regions is emphasized as the criterion of the classification model, the binary classification is the best as it can induce three regions with the highest coverage 100%. There are different rough set models using different mechanisms to obtain various three-way regions. The decision-theoretic rough set model (DTRS) obtains three regions by minimizing overall classification cost [37]. Based on DTRS, neighborhood based decision-theoretic rough set model is proposed to deal with numerical data with noise [18], multi-objective optimization method is adopted to learn the thresholds in DTRS [21]. In information-theoretic rough set model (ITRS), the regions are calculated by minimizing the sum of Shannon entropies of three regions which is one of measurements of uncertainty [6]. Game-theoretic rough set model (GTRS) obtains region thresholds by formulating competition or cooperation among multiple criteria or measures [7,15,29,30]. The criteria for evaluating three-way regions, such as accuracy, generality, or uncertainty, arrive at a compromise by competing with each other in games [1].

Different models use different measures to determine three-way regions. There are some measurements or criteria in other fields that we may utilize to evaluate regions. Gini coefficients may be used to examine the impurity of three-way regions [43,45,46]. The relationships and associations between regions' sizes and their impurities are studied in this paper. The changes of the three regions' sizes may be associated with the fluctuation of impurities. In other words, impurity may be used to configure three-way regions. Intuitively speaking, when objects are moved from acceptance or rejection region to non-commitment region, the impurity of acceptance or rejection region decreases and that of non-commitment region increases. When objects are moved from non-commitment region to acceptance or rejection region, the impurity of non-commitment region decreases and that of acceptance or rejection region increases. A preferred situation is that three regions have lower impurities. However, the impurities of regions are conflicting to each other. The decrease of one region's impurity is always accompanied by the increase of the other regions' impurities. We construct Gini objective functions to balance the conflict relationships among impurities of regions. Balanced three-way regions may be derived from these Gini objective functions. In particular, three different Gini objective functions, i.e., minimizing the summation of impurities, minimizing the impurities of immediate and non-commitment regions simultaneously, and minimizing three impurities simultaneously, are discussed in this paper. These Gini objective functions may be associated with specific requirements of different applications. We aim to find the suitable three-way regions which provide a trade-off among impurities of regions. The result in this study may enhance our understanding of three-way classifications and rough sets to make them practical in applications.

2. Three-way regions and related evaluations

In this section, we briefly introduce the background concepts used in this research, namely, the three-way regions formulated with rough sets and the measures for evaluating these regions.

2.1. Three-way regions formulated with rough sets

The three-way classifications are outlined by Yao in [34,35]. The universe of objects U is divided into acceptance, rejection and non-commitment regions based on the criteria C. The evaluation function and the specified acceptance and rejection thresholds are two issues that must be considered when formulating three-way classifications [34]. An evaluation function determines the particular region that an object belongs to [34]. If the evaluation value of an object is greater than the specified level for acceptance, we classify the object into an acceptance region. If the evaluation value of an object is lower than the specified level for rejection, we classify the object into a rejection region. If the evaluation value of an object is between the specified levels for acceptance and rejection, we classify the object into a non-commitment region. The rules of acceptance and rejection decisions can be induced from the acceptance and rejection regions, respectively [42]. For the objects in the non-commitment region, we may take a more cautious approach by not making a commitment [32]. Three-way classifications can be formulated from rough sets, fuzzy sets, shadow sets, interval sets, etc. [38,40,41]. These models provide different definitions of evaluation functions, as well as determinations and interpretations of specified thresholds for acceptance and rejection [39].

We formulate three-way classifications by using probabilistic rough sets. Suppose that the universe of objects U is a finite nonempty set. Let $E \subseteq U \times U$ be an equivalence relation on U, where E is reflexive, symmetric, and transitive [23]. For an element $x \in U$, the equivalence class containing x is given by $[x] = \{y \in U | xEy\}$. The family of all equivalence classes defines a partition of the universe U and is denoted by $U/E = \{[x] | x \in U\}$, that is, the intersection of any two elements is an empty set and the union of all elements are the universe U [23]. For an indescribable target concept $C \subseteq U$, probabilistic rough sets utilize conditional probability and thresholds (α, β) to define three-way regions, i.e., acceptance, rejection and non-commitment regions of C [31]: