Available online at www.sciencedirect.com
ScienceDirect
Computers and Mathematics with Applications 55 (2008) 1754–1765
An International Journal
computers & mathematics with applications
www.elsevier.com/locate/camwa

Converse approximation and rule extraction from decision tables in rough set theory

Yuhua Qian[a,b,c], Jiye Liang[a,b,*], Chuangyin Dang[c]

[a] *Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan, 030006, China*
[b] *School of Computer and Information Technology, Shanxi University, Taiyuan, 030006, China*
[c] *Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Hong Kong*

Received 14 November 2006; received in revised form 14 August 2007; accepted 30 August 2007

Abstract

In this paper, the concept of a granulation order is proposed in an information system. The converse approximation of a target concept under a granulation order is defined and some of its important properties are obtained, which can be used to characterize the structure of a set approximation. For a subset of the universe in an information system, its converge degree is monotonously increasing under a granulation order. This means that a proper family of granulations can be chosen for a target concept approximation according to user requirements. As an application of the converse approximation, an algorithm based on the converse approximation called REBCA is designed for decision-rule extraction from a decision table, which has a time complexity of $O(\frac{m}{2}|C|^2|U|\log_2|U|)$, and its practical applications are illustrated by two examples.

Keywords: Rough set theory; Converse approximation; Decision table; Decision rules; Rule extracting

1. Introduction

Recently, rough set theory developed by Pawlak in [1] has become a popular mathematical framework for the analysis of vague descriptions of objects. The focus of rough set theory is on the ambiguity caused by limited discernibility of objects in the domain of discourse. Its key concepts are those of object indiscernibility and set approximation, and its main perspectives are information view and algebra view [2]. The primary use of rough set theory has so far mainly been in generating logical rules for classification and prediction using information granules; thereby making it a prospective tool for pattern recognition, image processing, feature selection, data mining and knowledge discovery process from large data sets [3–5].

As a recently renewed research topic, granular computing (GrC) is an umbrella term to cover any theories, methodologies, techniques, and tools that make use of granules in problem solving [6–8]. Basic ideas of GrC have appeared in related fields, such as interval analysis, rough set theory, cluster analysis, machine learning,

* Corresponding author at: Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan, 030006, China.

E-mail addresses: jinchengqyh@126.com (Y. Qian), ljy@sxu.edu.cn (J. Liang), mecdang@cityu.edu.hk (C. Dang).

0898-1221/$ - see front matter
doi:10.1016/j.camwa.2007.08.031

databases, and many other. As follows, for our further development, we briefly review research on GrC. L.A. Zadeh identified three basic concepts that underlie the process of human cognition, namely, granulation, organization, and causation. "Granulation involves decomposition of whole into parts, organization involves integration of parts into whole, and causation involves association of causes and effects". A general framework of granular computing was presented by Zadeh [6] in the context of fuzzy set theory. Some authors [9–11] examined granular computing in connection with the theory of rough sets. Yao [12] suggested the use of hierarchical granulations for the study of stratified rough set approximations. Lin [13] and Yao [14] studied granular computing using neighborhood systems. Klir [15] investigated some basic issues of computing with granular computing with granular probabilities. The theory of quotient space had been extended into the theory of fuzzy quotient space based on fuzzy equivalence relation [16]. Liang and Shi [17] established the relationship among knowledge granulation, information entropy, granularity measure and rough entropy in rough set theory. Liang and Qian [18,19] studied rough sets approximation based on dynamic granulation (positive approximation) and its application for rule extracting. In the view of granular computing, a general concept described by a set is always characterized via the so-called upper and lower approximations under static granulation in rough set theory, and a static boundary region of the concept is induced by the upper and lower approximations. However, in rough sets approximation under dynamic granulation, a general concept described by using the positive approximation is characterized via variational upper and lower approximations under dynamic granulation, which is an aspect of people's comprehensive solving ability at some different granulations.

In recent years, rough set theory has been widely applied to extracted from decision tables. Decision tables have two cases: consistent decision tables and inconsistent decision tables. Ziarko [20] proposed a method of rule extracting from decision tables by reducing boundary area in decision tables. There are at least three approaches to reducing boundary area. The first and simplest technique is to try to increase the decision table "resolution" by adding more attributes or by increasing the precision of existing ones. The second approach is to provide another layer of decision tables, by treating each subdomain of objects matching the description of an elementary set of the boundary area of the original decision tables as a domain (the universe) by itself. The third proposed method of boundary area reduction is based on the idea of treating the subdomain of the original domain corresponding to the whole boundary area as the new domain by itself. However, in fact, the classification accuracy (the approximation measure) [21] is constrained according to decision requirements or preference of decision makers in general. An obvious question is how to extract much simpler decision rules on the basis of keeping an approximation measure. Liang and Qian [18,19] presented the notion of positive approximation and applied it for rule extracting from consistent decision tables in rough set theory. In [21], a given relative knowledge reduction determines a family of decision rules for a decision table. That is to say, relative knowledge reduction must be obtained before rule extracting from decision tables. Many types of knowledge reductions have been proposed [22–33] in the area of rough set and each of the reductions aimed at some basic requirement. However, the complexity of these attribute reductions are much worse, which is inconvenient to extract decision rules from decision tables. Our research aims to find a method for rule extracting without computing relative attribute reduction of a decision table in rough set theory. Based on these studies, the main objectives of this paper are to establish the structure of the approximation of a target concept by introducing a notion of a granulation order, investigate some of its important properties, and apply it to rule extracting from decision tables.

The rest of this paper is organized as follows. In Section 2, we review some basic concepts and properties of the positive approximation. The definitions of converse approximations of a target concept and a target decision (target partition) based on dynamic granulation are presented respectively, and some of their very useful properties are educed in Section 3. In Section 4, a new rule-extracting method of decision tables based on converse approximation in rough set theory is proposed and time complexity of this algorithm is analysed. And we show how the algorithm to extract decision rules by two illustrative examples (a consistent decision table and an inconsistent decision table). Finally, Section 5 concludes the whole paper.

2. Positive approximation

In this section we briefly introduce some existing basic concepts and properties of the positive approximation that are relative to the research in this paper. These concepts and properties will be helpful for us to understand the notion of a granulation order and set approximation under dynamic granulation and to establish converse approximation in this paper.