Information Sciences 381 (2017) 193–208

Contents lists available at ScienceDirect

Information Sciences

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

A new approach of optimal scale selection to multi-scale decision tables

Feng Li[a,b], Bao Qing Hu[a,b,*]

[a] School of Mathematics and Statistics, Wuhan University, Wuhan, 430072, PR China
[b] Computational Science Hubei Key Laboratory, Wuhan University, Wuhan, 430072, PR China

ARTICLE INFO

Article history:
Received 30 September 2015
Revised 20 September 2016
Accepted 25 November 2016
Available online 27 November 2016

Keywords:
Granular computing
Multi-scale decision tables
Optimal scale combination
Rough sets

ABSTRACT

As a special case of information table, multi-scale decision table can usually be observed in real-life world. In such table, objects may take different values under the same attribute measured at different scales. Based on inclusion relation of subsets of attributes and coarse relation of scales of attributes, multi-layered granulations and stratified rough set approximations in multi-scale decision tables are shown from the perspective of granular computing. Compared with a special case studied by Wu and Leung, the multi-scale decision tables of diverse attributes with different numbers of levels of scales are studied in this paper. Furthermore, complement model and lattice model are proposed to analyze the optimal scale selection for multi-scale decision tables. Correspondingly, algorithms of the two models are designed and some experiments are performed to testify feasibilities of these proposed algorithms and to make comparisons of the models.

1. Introduction

Since the topic of fuzzy information granulation has been firstly proposed and discussed by Zadeh in 1979 [60,61], researchers have a growing interest in the thought of information granulation. Granules may be understood as groups, classes, or clusters of elements, which are drawn together by indistinguishability, similarity, proximity, or functionality [61]. Granular computing (GrC), serving as an effective tool of complex problem solving, massive dataset mining and fuzzy information processing, has become a very active field of research in recent years [1,2,20,29,31,35,39,40,43,44,48,51,54–59].

GrC should be considered when it is impossible to distinguish the elements of a universe, which may be caused by incomplete, uncertain, or vague information involved in a system [54]. GrC has been widely applied in many fields. Lin [29,30] and Yao [53,54] studied GrC based on neighborhood systems to interpret granules. Partition model proposed by Yao [57] is an important and commonly used one for GrC, which is constructed by granulating a finite universe of discourse through a family of pairwise disjoin subsets under an equivalence relation. Dick et al. [8] and Zhou et al. [63] studied the neural network from the view of GrC. Zhang and Miao [62] investigated double-quantification approximation on GrC and rough set models. Belohlavek et al. [3], Li et al. [24,25] and Xu et al. [50] studied the combination of GrC with concept learning.

Rough set theory originally proposed by Palwak [33] has played a vital role in the extension and development of GrC. As a special tool of soft computing, it is able to model and process uncertainty or incomplete information in an intelligent

* Corresponding author. Fax:+86 27 68752256.
E-mail addresses: tsingfeng@whu.edu.cn (F. Li), bqhu@whu.edu.cn (B.Q. Hu).

http://dx.doi.org/10.1016/j.ins.2016.11.016
0020-0255/

system. And it performs well in construction, interpretation and representation of granules in a universe by an equivalence relation [47]. From the viewpoint GrC, equivalence granules can be constructed in Pawlak approximation space depending on an equivalence relation. And the elements in an equivalence granule are in some sense indistinguishable or similar. Thus, the equivalence granules are the basis for representation and approximation in Pawlak approximation space.

Attribute-value representation model is called an information table (sometimes called information system, data table etc.) [34], through which knowledge representation in rough set theory is realized. An information table that takes only one value for each object at each attribute is called a single-scale information table (SSIT). In a SSIT, an arbitrary subset of attributes determines an equivalence relation, on which granules are constructed. And the inclusion relation between subsets of attributes implies a coarse or fine relation of granules, which induces a multi-layered granulation structure on the universe. However, one may obtain hierarchically structured data in real-life world, in other words, there are different values under the same attribute for the same object when it is measured at different scales, especially, in the area of map/geographical information systems represented in multiple scales and remotely sensed data obtained at multiple resolutions. For example, Harvard University is located at Cambridge[1] in a finer granule compared to Boston University, or at Massachusetts in a fine granule compared to Yale University, or at New England[2] in a coarse granule compared to Duke University, or at the United States in a coarser granule compared to University of Oxford. In [46,47], such information tables are called multi-scale information tables (MSITs). For a given subset of attributes, two different levels of scales may induce a kind of granules being either a refinement or a coarsening of the others, which is the main difference between SSIT and MSIT. In general, finer granules cost more, thus an appropriate level of granulation should be selected to approximate subsets of a universe.

Multi-granulation rough set model (MGRS) is proposed by Qian et al. and extended Pawlak's rough set model [36–38]. Recently, the multigranulation approach attracts more and more researchers [16–19,27,28,52]. Furthermore, multi-scale information systems can be explained as an application of Model RI in a multigranulation space in [52], whereas it has some differences from a multigranulation rough set, which is a Model A in [52]. One is that MGSR is a methodology while MSIT is a special application of knowledge representation. Another is that MGSR is based on a family of approximations from a set of equivalence relations in a multigranulation space, while the models on MSIT in this paper are to select an approximation space of a multigranulation space for approximating better.

In an MSIT, values observed at different scales represent different knowledge with different granules and have different acquiring costs. Discarding some needless details and concentrating on a special level of detail are the key approaches for hierarchically structured information tables. Namely, we usually consider and analyze knowledge at the optimum level of granularity [47]. In [46], Wu and Leung introduced the notion of multi-scale decision tables (MSDTs) from the perspective of GrC and analyzed the knowledge acquisition in MSDTs under different levels of granulations. In [47], Wu and Leung mainly studied optimal scale selection for multi-scale decision tables with an assumption that each attribute is granulated with the same number of levels of granulations. On the same assumption, Gu et al. [12,13] and She et al. [41] studied the knowledge acquisitions and rule induction in MSDTs. However, the assumption may bring some restrictions on the applications of MSIT in real-life world. In this paper, we make a generalization that attributes may have different numbers of levels of granulations. Since various rough set models have been proposed for knowledge reduction and decision rules acquisition in information tables [4,7,9–11,15,21,23,32,42,45,49], optimal scale selection of the general multi-scale information systems is mainly studied in this paper. Besides, we only discuss consistent decision tables since an inconsistent decision table can be transformed into a consistent one.

For optimal scale selection in the special MSDTs, the approach introduced by Wu and Leung in [47] is called Wu–Leung model and in this paper, we propose complement model and lattice model for the general case. Then the corresponding algorithms are designed, and some experiments are employed to illustrate performances of the proposed models.

The other parts of the paper are organized as follows. In Section 2, several basic notions of Pawlak's rough set, information system, stratified rough set and multi-scale information system are reviewed. In Section 3, optimal scale selection of Wu–Leung model and our two new models are introduced. Algorithms for computing optimal scale selection via the new models and some experiments testing the algorithms are gave in Section 4. Finally, we conclude the paper with a summary and an outlook of further work in Section 5.

2. Preliminaries

In this section, we review some basic notions of Pawlak's rough set, information system, stratified rough set and multi-scale information system.

2.1. Pawlak's rough sets

Let U be a nonempty finite set of objects called the universe of discourse, and R an equivalence relation on U called an indiscernibility relation. Equivalence relation R is available knowledge for objects in the universe, and pair $apr = (U, R)$ is called a Pawlak approximation space. Universe U can be partitioned into disjoint pairwise subsets by R, and the collection

[1] Cambridge is a city in Massachusetts, United States.

[2] New England is a region which comprises six states of the Northeastern United States: Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont.