Knowledge-Based Systems 102 (2016) 103–115

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A variable precision rough set model based on the granularity of tolerance relation

Xiangping Kang [a,b,*], Duoqian Miao [a,b]

[a] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
[b] Key Laboratory of Embedded System and Service Computing (Tongji University), Ministry of Education, Shanghai 201804, China

ARTICLE INFO

Article history:
Received 10 December 2015
Revised 25 February 2016
Accepted 30 March 2016
Available online 1 April 2016

Keywords:
Rough set
Concept lattice
Tolerance relation
Tolerance class

ABSTRACT

As one of core problems in rough set theory, normally, classification analysis requires that "all" rather than "most"elements in one class are similar to each other. Nevertheless, the situation is just opposite to that in many actual applications. This means users actually just require "most" rather than "all"elements in a class are similar to each other. In the case, to further enhance the robustness and generalization ability of rough set based on tolerance relation, this paper, with concept lattice as theoretical foundation, presents a variable precision rough set model based on the granularity of tolerance relation, in which users can flexibly adjust parameters so as to meet the actual needs. The so-called relation granularity means that the tolerance relation can be decomposed into several strongly connected sub-relations and several weakly connected sub-relations. In essence, classes defined by people usually correspond to strongly connected sub-relations, but classes defined in the paper always correspond to weakly connected sub-relations. In the paper, an algebraic structure can be inferred from an information system, which can organize all hidden covers or partitions in the form of lattice structure. In addition, solutions to the problems are studied, such as reduction, core and dependency. In short, the paper offers a new idea for the expansion of classical rough set models from the perspective of concept lattice.

1. Introduction

It is known that concept has been taken as the unit or cell of human cognition in people's thinking activities, since concept contains the most essential information of some kind of things, it plays an important role in human's cognitive process. In essence, as one major method for human to know the real world and its laws, concept thinking can be served as the foundation for people to form various complicated ideas and also effective means to express knowledge. In 1982, German mathematician Wille professor brought forth formal concept analysis (FCA), or concept lattice theory [46], which can be considered as an application branch of lattice theory. As one kind of method to mathematically abstract and formalize concepts from the objective world, FCA greatly stimulates people's enthusiasm to solve problems under the concept thinking. In FCA, the basic viewpoint of concept essentially developed from the understanding of concept in philosophy, that is, one concept is mathematically described from aspects of extent and intent, in which extent refers to the set of objects covered by concept, and intent refers to the set of common characteristics of objects covered by concept. Concept lattice, as the core data structure of FCA, is an effective tool for data analysis and rule extraction, and can vividly and concisely manifest the generalization-specialization relationship among concepts by means of Hasse graph. In recent years, concept lattice has developed into a powerful data analysis method [14,16,21,25,32,37,48,52], and found wide applications in many fields like data mining analysis, information retrieval, knowledge discovery, ontology engineering, etc.

In practice, information collected from actual systems often contains noise, namely, information is not always accurate or complete. Along with the rapid development of science and technology, the uncertainty of information is more and more remarkable. Therefore, it is always inevitable for people to process the uncertainty and incomplete information in various applications. In the case, how to distill useful knowledge from the massive, inaccurate, fuzzy or incomplete information has become an extremely urgent task. Although, people can use pure mathematical assumptions to eliminate or avoid this uncertainty, but the effect is often not ideal. Conversely, if methods can appropriate to deal with these information, it is often helpful to solve many complex practical problems. Over the years, researchers have been trying to find effective ways to deal with the incomplete and uncertainty information

* Corresponding author. Tel.: +86 0351 6563220.
E-mail addresses: tongji_kangxp@sina.com (X. Kang), miaoduoqian@163.com (D. Miao).

http://dx.doi.org/10.1016/j.knosys.2016.03.030
0950-7051/

scientifically. As classic methods to deal with uncertain information, evidence theory, fuzzy theory, probability statistics, etc. have been used in many practical fields. However, these methods need some additional information or prior knowledge, such as fuzzy membership function, belief function, statistical distribution function, etc. which can not be easily obtained.

In 1982, Polish scholar Pawlak brought forth rough set theory [31], as a kind of important reasoning technology in artificial intelligence, which can effectively analysis and process the fuzzy and uncertain information without any prior knowledge except for data sets. Its main idea is to, with the classification ability being kept unchanged, deduce decision or classification rules of problems through knowledge reduction. Meanwhile, it can use the observed and measured knowledge to approximately describe imprecise or uncertain concepts. Due to its effectiveness and usability in the process of dealing with uncertain problems, rough set has already drawn much attention of scholars [9,11,12,33], and lots of research results have been widely applied to various fields, such as medical diagnosis, decision analysis, image processing, machine learning, and so on. In addition, with the deepening research and widening scope, the data forms and organization structures are increasingly diversified, so it becomes more and more difficult for people to effectively solve the complicated practical problems just through any single theory. Therefore, combining rough set with other artificial intelligence technology has become a hot research topic of international scholars, such as probability statistics, fuzzy set, evidence theory, neural network, concept lattice [35,39,44,47], and so on. So far, the whole theoretical system of rough set has already been gradually maturing and increasingly perfect, which greatly enriched and expanded the theoretical foundation and the application scope of rough set.

Rough set and concept lattice, as two mathematical branches generated in the same era, there are some significant differences from the perspective of their research methodology, but the same research background and objective indicate that they must have something in common. In fact, the two theories share many similarities [17], such as any one-valued formal context, as a kind of data set, is just a special case of information systems essentially, therefore, their mutual reference and integration not only enhance their own analytic abilities, but also can help to understand one theory from the perspective of another. Meanwhile, by means of mixing their respective advantages, the fusion theory may help to establish a more general and universal data analysis framework. Therefore, it is extremely significant to combine two theories in terms of their advantages. Recently, many remarkable research achievements of the fusion theory have been made. Oosthuizen informally described the connection between rough set and concept lattice [30]. In the study of logical models, Duntsch and Gediga defined modal-style operators on the basis of binary relations, and constructed the attribute-oriented concept lattice according to the upper approximate operator [5,8]. Deogun and Saquer mainly discussed the monotone concept lattice, which is a direct expansion of classical concept lattice [4,36]. By introducing the idea of upper and lower approximations in rough set, Yao expended the definition of concept lattice, studied the rough set approximation of formal concept, built object-oriented and attribute-oriented concept lattices, and proved that attribute-oriented concept lattice and object-oriented concept lattice are isomorphic [49–51]. Zhang et al. introduced variable threshold concept lattices [53]. Belohlavek et al. provided the uniform structure of different variable threshold concept lattices [1]; Fan et al. studied fuzzy inferences based on fuzzy concept lattices [6]. Through comparing the relationship between fuzzy concept lattice and rough set, Lai et al. pointed out that each complete fuzzy concept lattice could be expressed as the concept lattice in the sense of rough set under certain conditions [18]. Lots of scholars introduced the idea of reduction in rough set into concept lattice, and discussed the reduction theory in concept lattice [2,20,24,27–29,45]. Kang et al. once suggested a rough set model based on concept lattice, which solved the problem of algebraic structure in the discrete information system, namely inducing a lattice structure from an information system, with each node in the lattice being called a rough concept, meanwhile, they also presented solutions to some common problems in rough set based on concept intents, such as core, reduction and function dependence [15]. For more flexible and efficient learning concept, from the cognitive computing perspective, Li et al. investigated concept learning by means of granular computing and set approximations [22], in addition, they have focused on issues of approximate concept lattice, approximate decision rule and knowledge reduction in incomplete decision contexts [23]. Shao and Leung revealed some relationships of reduction results in rough set and concept lattice [38]. Tan systematically explored connections between rough set and concept lattice in terms of approximation operators, structures and knowledge reduction [43]. Li et al. [26] made a comparison between multigranulation rough sets and concept lattices via rule acquisition, and obtained some interesting results. For more research findings concerning the fusion theory of rough set and concept lattice, please see the literature [51].

It is known to us that Pawlak's classical rough set model is established on the basis of equivalence relation (equivalence relation needs to meet reflexivity, transitivity and symmetry), and used to process complete information systems containing nominal attributes (domain of attribute is composed of several discrete values, and different values are independent of each other). However, when the domain of attribute is a real number set, or the differences among different values are caused by test errors, or the problems to be solved are highly complicated, or the scale of data set is too big, it is meaningless to analyze some minor differences. In the case, classical Pawlak's rough set model obviously has some limitations. In practical applications, users may not only require that objects with identical attribute values should be put into the same class, but also assume that objects with similar attribute values should also be classified the same. To further enhance the data processing capability of rough set, many scholars expand equivalence relations to tolerance relations (sometimes called similarity relation) only meeting reflexivity and symmetry. Tolerance relation is substantially different from binary relation of other types in terms of symmetry. Namely, symmetry is the basic characteristic of tolerance relation. In view of the universality of tolerance relation, great research findings have been made on the theory and application of rough set based on tolerance relation in recent years. To enhance the data processing capability of rough set, Slowinski et al. studied the properties and applications of rough set based on similarity relation, and pointed out that rough set based on similarity relation can be used for ignoring minor differences of attribute values [41,42]; Skowron et al. presented rough set based on tolerance relation, which was conducive to enhancing the robustness of system decisions and also the efficiency of decision making [40]; Leung and Li [19] studied the granules in incomplete information system, namely, with maximal tolerance classes as information granules, overcame the flaws of knowledge expression based on similarity class. Hu et al. proposed neighborhood rough set models in information systems with mixed features, where objects with numerical attributes were granulated with fuzzy tolerance relations obtain by Euclidean distance, while objects with nominal features were granulated with equivalence relations [13]. Guan and Wang applied maximal tolerance classes to set-valued information system, and discussed problems of attribute reduction and decision rule acquisition [10]. Based on maximal tolerance classes, Qian et al. studied the approximation reduction in inconsistent incomplete decision tables [34]. Dai defined fuzzy tolerance