Knowledge-Based Systems 105 (2016) 147–159

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A study on information granularity in formal concept analysis based on concept-bases

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 13 December 2015
Revised 18 March 2016
Accepted 7 May 2016
Available online 9 May 2016

Keywords:
Concept lattice
Granularity
Concept-bases
Concept similarity

ABSTRACT

As one of mature theories, formal concept analysis (FCA) possesses remarkable mathematical properties, but it may generate massive concepts and complicated lattice structure when dealing with large-scale data. With a view to the fact that granular computing (GrC) can significantly lower the difficulty by selecting larger and appropriate granulations when processing large-scale data or solving complicated problems, the paper introduces GrC into FCA, it not only helps to expand the extent and intent of classical concept, but also can effectively reduce the time complexity and space complexity of FCA in knowledge acquisition to some degree. In modeling, concept-base, as a kind of low-level knowledge, plays an important role in the whole process of information granularity. Based on concept-base, attribute granules, object granules and relation granules in formal contexts are studied. Meanwhile, supremum and infimum operations are introduced in the precess of information granularity, whose biggest distinction from traditional models is integrating the structural information of concept lattice. In addition, the paper also probes into reduction, core, and implication rules in granularity formal contexts. Theories and examples verify the reasonability and effectiveness of the conclusions drawn in the paper. In short, the paper not only can be viewed as an effective means for the expansion of FCA, but also is an attempt for the fusion study of the two theories.

1. Introduction

With the increasing popularization of internet technology, “rich data and scarce knowledge” has gradually become a more and more important problem. In the case, how to intelligently and automatically extract potential knowledge from the large-scale data has become one of research hotspots in the current data mining field. Essentially, data mining is a process from the data to information to knowledge, similarly, from the perspective of concept cognition, it can also be understood as the process from the data to lower concepts to higher concepts. Concept cognition is an important characteristic of human brain learning, which is a thinking pattern formed in our minds and mainly focuses on concepts formed through the abstraction and summarization of the common essential characteristics of things. In fact, as one of the most important ways to understand the real world and its regularity, concept cognition is an important foundation of people’s complicated thought, and it is also an effective means to express and deduce knowledge.

In philosophy, concept is the thinking unit of human understanding of the objective world and its law, which is composed of two parts, namely intent and extent. In this sense, concepts are essentially the abstract, generalization and induction of the objective world, and the process of generating concepts is essentially a process of optimization and evolution from the perceptual to the rational, and from the phenomenon to the essence, and from the scatter to the system. Meanwhile, in order to adapt to rapid changes in the subjective and objective world, concepts are not only the summary of the understanding of objective things, but also the starting point of the new knowledge, people can deduce new concepts from known ones. If “knowledge system” is compared to a building, then concepts can be understood as core elements of it. Therefore, developing new techniques and methods based on concept thinking will surely contribute to the rapid development of data mining.

If people can use methods in mathematical form to simulate the formation process of concepts and discover the relationship among concepts at different levels, then it will has great significance for data mining and knowledge discovery. Therefore, based

* Corresponding author.
E-mail addresses: tongji_kangxp@sina.com (X. Kang), tongji_miaodq@sina.com (D. Miao).

http://dx.doi.org/10.1016/j.knosys.2016.05.005
0950-7051/

on the philosophical understanding of concepts, German scholar Wille proposed FCA [36] in 1982 from the Brikhoff's lattice theory [1], its formal system can well describe the formation process of concepts by mathematical methods, and can help to stimulate people's mathematical thinking for data analysis and knowledge processing under the concept cognition. In FCA, concepts, concept lattice, Galois connection, et al. play greatly important roles. Namely, any concept can be characterized and described from perspectives of intent and extent, which helps to deepen people's accurate understanding and summary cognizance of actual concepts; Concept lattice, as the core data structure of FCA, can intuitively reflect the generalization and specialization relationships between concepts through Hasse graph; Galois connection is the core theoretical basis of concept lattice. In addition, the generating process of concept lattice is actually a process of objects clustering. In recent years, the research on concept lattice has made a series of important results [15,17,22,23,26,33,43]. For current trends and directions of concept lattice, please refer to reference [28].

In recent years, as a mature theory possessing solid mathematical properties, FCA has drawn more and more attention and been widely applied in various fields such as machine learning, decision analysis, data mining. However, along with the development of research, it is difficult to effectively solve complicated problems just through one single theory. Therefore, many scholars have combined FCA with other theories such as fuzzy set [2,8,12,25,42], rough set [3,9,11,14,21,31,34,35,37], neural network [4], probability theory [7], GrC [8,10,16,29,38], et al., thus greatly expanding the theoretical foundation and application scope of FCA. For example, Burusco et al. brought fuzzy theory into FCA, which can accurately express the uncertain relationship between attributes and objects, thus breaking the binary limitations of classical FCA and helping solve fuzzy and uncertain information in actual applications [2]; Dias et al. combined neural network and FCA and proposed the FCANN [4]; Jiang et al. combined probability theory and FCA, and proposed a new data mining method SPICE [7]; Kang et al. introduced GrC into FCA, and provided a unified model for concept lattice building and rule extraction on the basis of a fuzzy granularity base [8]; The reference [10] introduced FCA and GrC into ontology learning, and presented a unified research model for ontology building, ontology merging and ontology connection based on the domain ontology base in different granulations; Kent discussed the relationship between concept lattice and rough set theory, and presented rough concept analysis that can be viewed as a synthesis of rough set and FCA [11]; Tan et al. studied connections between covering-based rough sets and FCA [31]; Ventos and Soldano presented α Galois lattices based on equivalence classes [32].

FCA, as a data analysis tool, possesses characteristics such as completeness and precision. For a classical concept, even though an object possesses most of attributes of the intent, the object is still not included in the extent of the concept. At the time, it is difficult to manifest the object "possibly" included in concept. Such precision is an advantage of FCA, but it also results in some limitations in processing some specific knowledge. For example, in the process of earthquake prediction, the earth may not always has all characteristics of the earthquake, so experts can only judge that the earthquake possibly happen. However, if all characteristics are identical, then the earthquake may have come about. Since the consequence may be serious, the possibility can not be overlooked. In addition, in some large-scale or complicated data, as a matter of fact, it still deserves attention that such precision of concept lattice often results in a mass of concepts and makes the structure of concept lattice extremely complicated. Aiming at above problems, to better understand and solve problems rather than get lost in unnecessary details, and to better discover potentially valuable knowledge from seemingly irrelevant data, the paper introduces the idea of GrC into FCA, which can, via the unique advantage in the modeling and analysis of large-scale complicated data, lower the "resolution" of knowledge acquisition, and expand the "scale" of knowledge measurement, thus can effectively simplify the complicated concept lattice structure and compact the huge concept scale and prevent some useful information from being buried in massive information.

As a feasible and effective solution for data mining and knowledge reasoning, GrC has become a hotspot research subject in the artificial intelligence field, and a great deal of articles were published. American famous mathematician Zadeh [40] first presented and discussed fuzzy information granularity on the basis of the fuzzy set theory in 1979. In 1997, "granular computing" concept was first formally presented. Some important theoretical findings concerning GrC are shown as follows: Aiming to solve the problem of fuzzy intelligent control through natural language fuzzy reasoning and judgment, Zadeh presented the theory of computing with words [41], and then Thiele proposed the semantic models for investigating computing with words [30], which facilitated the development of the theory of computing with words; Pawlak proposed the rough set theory used for uncertain knowledge modeling, and offered a modeling tool in the case that priori knowledge was incomplete or uncertain [24]; Lin and Yao emphatically described the significance of GrC, that roused people's tremendous interest [18,39]; Zhang et al. presented the GrC model based on quotient space often used in solving complicated problems [44]; Leung and Li described the basic "granule" in an information system with maximal consistent blocks [19]; Hu studied mixed data-oriented neighborhood relation granular computing model and its application [6]; Liang, Qian et al. conducted the systematic research on various uncertain measures, axiomatization of granulation measurement, and rule acquisition in information systems, and studied problems such as incomplete multi-granulation rough set and granulation space structure [20,27].

GrC is a new theory effectively simulating human's thinking and solving complicated problems in the intelligent information processing field. It possesses the unique advantage in the modeling and analysis of large-scale complicated data. No matter from the macro-perspective of cognitive philosophy or from the micro-perspective of information processing, GrC characterized by information granulation, relationships between granules, and granule-based reasoning essentially reflects human's features in solving complicated problems. GrC changed some of our traditional computing concepts in actual applications, making it more scientific, reasonable and operable to deal with problems. For example, when problems are too complicated or solving them requires high cost, the method no longer focuses on some inessential detailed information and takes mathematical exact solutions as the goal, but replaces exact solutions with the feasible satisfactory approximate solutions on the basis of the actual needs so as to achieve the goals of simplifying the problems and enhancing the problem-solving efficiency.

Normally, there are two types of knowledge acquisition methods based on the fusion theory of FCA and GrC. Namely, one type is indirect, which needs data preprocessing, and further using traditional methods to acquire knowledge; While another type is direct [32], which does not need data preprocessing, and can directly deal with original formal contexts. In fact, the former is relatively simple and easy to use, while the latter can completely preserve the original information, which may be more objective than the former.

For instance, for the formal context shown in Table 1(a), indirect methods need data preprocessing, namely, transforming it into a granularity context like Table 1(b), and further using following classical operators (see Definition 1) to acquire knowledge in