Chapter 7
Granular Computing Based Data Mining in the Views of Rough Set and Fuzzy Set

Guoyin Wang
Chongqing University of Posts and Telecommunications, P.R. China

Jun Hu
Chongqing University of Posts and Telecommunications, P.R. China

Qinghua Zhang
Chongqing University of Posts and Telecommunications, P.R. China

Xianquan Liu
Chongqing University of Posts and Telecommunications, P.R. China

Jiaqing Zhou
Southwest Jiaotong University, P.R. China

ABSTRACT

Granular computing (GRC) is a label of theories, methodologies, techniques, and tools that make use of granules in the process of problem solving. The philosophy of granular computing has appeared in many fields, and it is likely playing a more and more important role in data mining. Rough set theory and fuzzy set theory, as two very important paradigms of granular computing, are often used to process vague information in data mining. In this chapter, based on the opinion of data is also a format for knowledge representation, a new understanding for data mining, domain-oriented data-driven data mining (3DM), is introduced at first. Its key idea is that data mining is a process of knowledge transformation. Then, the relationship of 3DM and GRC, especially from the view of rough set and fuzzy set, is discussed. Finally, some examples are used to illustrate how to solve real problems in data mining using granular computing. Combining rough set theory and fuzzy set theory, a flexible way for processing incomplete information systems is introduced firstly. Then, the uncertainty measure of

DOI: 10.4018/978-1-60566-324-1.ch007
coverings Based rough set is studied by converting a covering into a partition using an equivalence domain relation. Thirdly, a high efficient attribute reduction algorithm is developed by translating set operation of granules into logical operation of bit strings with bitmap technology. Finally, two rule generation algorithms are introduced, and experiment results show that the rule sets generated by these two algorithms are simpler than other similar algorithms.

1. INTRODUCTION

Across a wide variety of fields, data are being collected and accumulated at a dramatic pace. It is necessary to acquire useful knowledge from large quantity of data. Traditionally, data mining is considered as the nontrivial extraction of implicit, previously unknown, and potentially useful information from data. That is to say, knowledge is generated from data. But in our opinion, knowledge is originally existed in the data, but just not understandable for human. In a data mining process, knowledge existed in a database is transformed from data format into another human understandable format like rule.

Granular computing is a label of theories, methodologies, techniques, and tools that make use of granules in the process of problem solving. The philosophy of granular computing has appeared in many fields, and it is likely playing a more and more important role in data mining. Rough set theory and fuzzy set theory are two very important paradigms in granular computing.

In this chapter, a new understanding for data mining, domain-oriented data-driven data mining (3DM), will be proposed. Moreover, we will introduce basic concepts of granular computing and analyze the relationship of granular computing and 3DM. We will also discuss the granular computing based data mining in the views of rough set and fuzzy set, and introduce some applications of granular computing in data mining. In the end, some conclusions are drawn.

2. BASIC CONCEPTS OF RELATED THEORIES

Domain-Oriented Data-Driven Data Mining (3DM)

Data mining (also known as Knowledge Discovery in Databases - KDD) is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data (Frawley, Piatetsky-Shapiro, & Matheus, 1991). It uses machine learning, statistical and visualization techniques to discover knowledge from data and represent it in a form that is easily comprehensible for humans. There are many commonly used techniques in data mining like artificial neural networks, fuzzy sets, rough sets, decision trees, genetic algorithms, nearest neighbor method, statistics based rule induction, linear regression and linear predictive coding, et al.

Unfortunately, most data mining researchers pay much attention to technique problems for developing data mining models and methods, while little to basic issues of data mining. What is data mining? Our answer would be “data mining is a process of knowledge transformation”. It is consistent with the process of human knowledge understanding. In our opinion, data is also a format for knowledge representation. The knowledge we mined from data was originally stored in data. Unfortunately, we cannot read, understand, or use it, since we cannot understand data. In a data mining process, we are
Related Content

Load Balancing to Increase the Consistency of Replicas in Data Grids
www.igi-global.com/chapter/load-balancing-increase-consistency-replicas/64440?camid=4v1a

Assisting Efficient Job Planning and Scheduling in the Grid
www.igi-global.com/chapter/assisting-efficient-job-planning-scheduling/20505?camid=4v1a

An Approach for Detection of Overloaded Host to Consolidate Workload in Cloud Datacenter
www.igi-global.com/article/an-approach-for-detection-of-overloaded-host-to-consolidate-workload-in-cloud-datacenter/202412?camid=4v1a

Structural Outlooks for the OTIS-Arrangement Network
Ahmad Awwad, Jehad Al-Sadi, Bassam Haddad and Ahmad Kayed (2013). Applications and Developments in Grid, Cloud, and High Performance Computing (pp. 221-231).
www.igi-global.com/chapter/structural-outlooks-otis-arrangement-network/69037?camid=4v1a