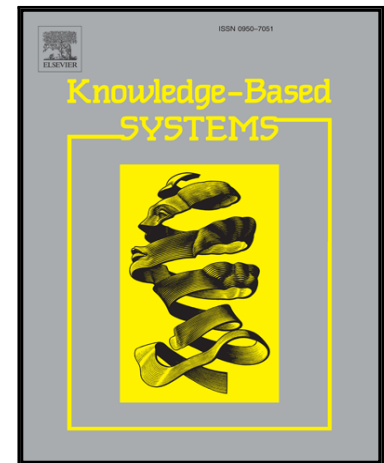


Attribute reduction of covering decision systems by hypergraph model

Jinkun Chen, Yaojin Lin, Guoping Lin, Jinjin Li, Yanlan Zhang

PII: S0950-7051(16)30446-4
DOI: [10.1016/j.knosys.2016.11.010](https://doi.org/10.1016/j.knosys.2016.11.010)
Reference: KNOSYS 3734



To appear in: *Knowledge-Based Systems*

Received date: 1 March 2016
Revised date: 11 November 2016
Accepted date: 13 November 2016

Please cite this article as: Jinkun Chen, Yaojin Lin, Guoping Lin, Jinjin Li, Yanlan Zhang, Attribute reduction of covering decision systems by hypergraph model, *Knowledge-Based Systems* (2016), doi: [10.1016/j.knosys.2016.11.010](https://doi.org/10.1016/j.knosys.2016.11.010)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Attribute reduction of covering decision systems by hypergraph model

Jinkun Chen^{a,*}, Yaojin Lin^b, Guoping Lin^a, Jinjin Li^a, Yanlan Zhang^b

^a*School of Mathematics and Statistics,
Minnan Normal University, Zhangzhou 363000, P. R. China*

^b*School of Computer Science,
Minnan Normal University, Zhangzhou 363000, P. R. China*

Abstract

Attribute reduction (also called feature subset selection) plays an important role in rough set theory. Different from the classical attribute reduction algorithms, the methods of attribute reduction based on covering rough sets appear to be suitable for numerical data. However, it is time-consuming in dealing with the large-scale data. In this paper, we study the problem of attribute reduction of covering decision systems based on graph theory. First, we translate this problem into a graph model and show that finding the attribute reduction of a covering decision system is equivalent to finding the minimal vertex cover of a derivative hypergraph. Then, based on the proposed model, a new attribute reduction algorithm for covering decision systems is presented. Experiments show that the new proposed method is more effective to handle the large-scale data.

Keywords: Attribute reduction, Decision systems, Rough sets, Hypergraphs, Vertex covers.

1. Introduction

Attribute reduction (feature selection) is considered to be a fundamental preprocessing step in pattern recognition and machine learning tasks. In classification learning, data are usually represented by an information system of objects that are described by a collection of features. However, many of these features are irrelevant or redundant which make the recognition problem more difficult. Thus, reducing the number of irrelevant and redundant features by selecting only the relevant ones is in demand. The main goal of attribute reduction is to remove redundant information in a data set, which can reduce not only the runtime of a learning algorithm, but in some cases it can also provide a better classification accuracy. Various methods have been developed for this purpose. Rough set theory, proposed by Pawlak, is a useful mathematical tool for feature selection [40, 41, 42]. It is also widely discussed and used in rule extraction, data mining and knowledge discovery [3, 8, 18, 22, 24, 25, 26, 27, 28, 29, 30, 44, 45, 46, 47, 48, 49, 53, 56, 59, 60, 62, 64].

In the classical rough set, the reduction is to find a minimal attribute set that provides the same approximating power as the whole set of attributes. With this definition, most of the irrelevant and redundant attributes can be removed. Many approaches have been developed to find the attribute reduction in the literature [19, 21, 23, 33, 34, 35, 43, 50, 51, 55, 65, 66]. For example, Qian et al. proposed four accelerating algorithms to find the attribute reduction [50]. However, the methods for attribute reduction with the classical rough set are only suitable for the categorical data and can not be directly applied to the numerical data. In other words, the numerical data need to be discretized before attribute reduction, and this causes information loss [20, 57]. For this reason, some generalized models were developed, such as neighborhood rough sets, similarity relation rough sets and covering rough sets [4, 5, 10, 12, 15, 16, 17, 54, 63]. Neighborhood rough sets have been proven to be a nice generalized model. In [15], Hu et al. proposed an effective feature selection method for the heterogeneous data. Recently, covering rough sets are used to study the attribute reduction. Chen et al. [6] proposed a new theoretical framework with covering rough sets

*Corresponding author.

Email addresses: cjk99@163.com (Jinkun Chen), yjlin@mnnu.edu.cn (Yaojin Lin)

Download English Version:

<https://daneshyari.com/en/article/4946379>

Download Persian Version:

<https://daneshyari.com/article/4946379>

[Daneshyari.com](https://daneshyari.com)