

工程与应用

基于粗糙集和聚类的纳税属性离散化方法

徐林章, 韩 臻, 张艳宁

西北工业大学 计算机学院, 西安 710072

收稿日期 2008-3-19 修回日期 2008-6-12 网络版发布日期 2008-10-8 接受日期

摘要 将粗糙集理论中属性重要度和依赖度的概念与分级聚类离散化算法相结合, 提出了一种纳税人连续型属性动态的离散化算法。首先将纳税数据对象的每个连续型属性划分为2类, 然后利用粗糙集理论计算每个条件属性对于决策属性的重要度, 再通过重要度由大至小排序进行增类运算, 最后将保持与原有数据对象集依赖度一致的分类结果输出。该算法能够动态地对数据对象进行类别划分, 实现纳税人连续型属性的离散化。通过采用专家分析和关联分析的实验结果, 验证了该算法具有较高的纳税人连续型属性离散化精度和性能。

关键词 [粗糙集](#) [分级聚类](#) [离散化](#) [数据预处理](#) [税源分析](#)

分类号

Discretization method of taxpayer' s continuous attribute based on rough set and cluster analysis

XU Lin-zhang, HAN Zhen, ZHANG Yan-ning

Department of Computer Science and Engineering, Northwest Polytechnical University, Xi' an 710072, China

Abstract

A dynamic discrete algorithm method to taxpayers' continuous attributes is proposed through combining the conception of attribute importance degree as well as dependence degree in rough sets and clustering discrete algorithm method. First, categorize each continuous attribute of the tax data objects into 2 groups, and then calculate the importance degrees of each conditional attribute related to its decision attribute. Second, make adding-categories calculations through sorting the importance degrees in ascending order. Last, output those classification results which have the same dependence degrees with original data objects set. This algorithm method can classify the data objects dynamically and realize the discretization of taxpayers' continuous attributes. This algorithm method has been tested in the experimental results of expert analysis and association analysis, which is proven that this method has higher accuracy and property of taxpayers' continuous attributes discretization.

Key words [rough set](#) [clustering](#) [discrete](#) [data pre-processing](#) [tax source analysis](#)

DOI: 10.3778/j.issn.1002-8331.2008.29.068

通讯作者 徐林章 xlz0088@sina.com

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(497KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [参考文献](#)

服务与反馈

- ▶ [把本文推荐给朋友](#)
- ▶ [加入我的书架](#)
- ▶ [加入引用管理器](#)
- ▶ [复制索引](#)
- ▶ [Email Alert](#)
- ▶ [文章反馈](#)
- ▶ [浏览反馈信息](#)

相关信息

- ▶ 本刊中 [包含“粗糙集”的相关文章](#)
- ▶ 本文作者相关文章

- [徐林章](#)
- [韩 臻](#)
- [张艳宁](#)