

数据库、信号与信息处理

基于粗糙集分类算法研究与实现

李勃¹, 王艳兵², 姚青²

1. 鲁东大学 计算机科学与技术学院, 山东 烟台 264025

2. 山东大学 计算机科学与技术学院, 济南 250061

收稿日期 2007-9-3 修回日期 2007-11-23 网络版发布日期 2008-5-16 接受日期

摘要 数据挖掘是人工智能中知识发现的重要组成部分, 而分类又是一种主要的形式。ID3算法是数据挖掘中经典的决策树分类算法, ID3算法具有抗噪声能力差的缺点。通过对分类和粗糙集理论的研究, 将可变精度粗糙集理论的思想应用在计算属性信息熵时设定阈值上, 以放宽属性选择的要求, 从而对经典的ID3算法作了相应的改进。改进后的ID3算法(称之为VPID3算法)可在一定程度上降低噪声对系统分类的干扰, 提高了有数据有噪声情况下的分类精度。另外根据该算法设计并实现了一个分类器, 并通过实验检验了该算法的性能。

关键词 [数据挖掘](#) [分类](#) [决策树](#) [粗糙集](#) [ID3](#) [熵](#)

分类号

Research and realization of classification based on rough set theory

LI Bo¹, WANG Yan-bing², YAO Qing²

1. School of Computer Science and Technology, Ludong University, Yantai, Shandong 264025, China

2. School of Computer Science and Technology, Shandong University, Ji'nan 250061, China

Abstract

Data mining is an important part of AI and classification is a kind of useful application. ID3 algorithm is a classical algorithm in data mining, the algorithm has the worse ability to resist noise. Through the research on variable precision rough set, the algorithm is improved by setting threshold value while calculating attributes' entropy, in order to relax the restrictions while selecting attributes. After using the improved ID3 algorithm (VPID3), the interference of noise to classification could be reduced to a certain extent, this made result correspond to reality even more. Finally, the paper designs and realizes a classifier using VPID3 algorithm and do some experiments to check its performance. Extensive experiments with four different datasets have shown that our algorithm is more effective in dealing with noise data than ID3 algorithm.

Key words [data mining](#) [classification](#) [decision tree](#) [rough set](#) [ID3](#) [entropy](#)

DOI:

通讯作者 李勃

扩展功能**本文信息**

- ▶ [Supporting info](#)
- ▶ [PDF\(570KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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“数据挖掘”的相关文章](#)
- ▶ [本文作者相关文章](#)

- [李勃](#)
- [王艳兵](#)
- [姚青](#)