数据库与信息处理

基于谱系聚类的粗糙集数据挖掘预处理方法

韩中华,马 斌,许 可,李宏亮

沈阳建筑大学 信息与控制工程学院，沈阳 110168

摘要　　介绍了一种基于统计分析的数据离散化方法——谱系聚类法，以胶合板缺陷检测数据为应用对象进行了基于谱系聚类的数据离散化研究，并与其它离散化方法进行了对比分析，对比结果表明经谱系聚类方法离散化后的数据，再进行粗糙集约简时，会有更多的冗余属性和记录被约掉，从而可以降低模型的复杂程度，加快获取知识的进程，提高分类的准确率。工程实践证明谱系聚类是一种有效的可用于数据预处理的离散化方法，结合粗糙集算法可以获取满意的数据挖掘结果。

关键词　　粗糙集　离散化　谱系聚类　类平均距离　SAS

分类号

Preprocessing algorithm based on pedigree cluster for rough set data mining

HAN Zhong-hua,MA Bin,XU Ke,LI Hong-liang

Faculty of Information and Control Engineering，Shenyang Jianzhu University，Shenyang 110168，China

Abstract

A data decentralize method based on statistical analysis is introduced，which is called pedigree cluster method，the data decentralize research，in which the inspection data of wood veneer are taken as the application，has been done and the comparison between pedigree cluster and several other decentralize method are also made.The comparison result shows that the data handled by pedigree cluster will be deleted more redundant attributes and records after rough sets theory reduction，the complexity of the model can be reduced，the knowledge acquisition process can be accelerated and the accurate of classification can be improved.It has been proved in engineering practice that pedigree cluster method is an effective data decentralize method used in preprocessing process，and combing with rough set method a satisfied data mining result can be obtained.

Key words　 rough sets　decentralize　pedigree clusters　group average distance　SAS

DOI:

通讯作者　韩中华 xiaozhonghua197709@163.com

扩展功能

本文信息

▶ Supporting info
▶ PDF(603KB)
▶ [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

· 韩中华
· 马 斌
· 许 可
· 李宏亮