

数据库与信息处理

分类数据集的一致化特征选择约简

吴新玲^{1,2}

1.广东技术师范学院 信息工程系,广州 510262

2.武汉大学 软件工程国家重点实验室,武汉 430072

收稿日期 修回日期 网络版发布日期 2007-6-9 接受日期

摘要 样本数据集的不一致性和冗余特征会降低分类的质量和效率。提出了一种一致化特征选择约简方法,该方法基于贝叶斯公式,采用阈值,将非一致数据归为最可能的一类,使数据集一致化。并在一致数据集上,运用类别区分矩阵选择可准确区分各类数据的最小特征变量集。给出的启发式搜索策略和应用实例表明:一致化特征选择约简方法能有效消除分类数据集的不一致性,选择最优的特征变量、降低数据的维数、减少数据集中的冗余信息。

关键词 [数据挖掘](#) [分类](#) [特征选择](#) [数据约简](#)

分类号

Consistent feature selection reduction about classification data set

WU Xin-ling^{1,2}

1.Department of Information Engineering,Guangdong Polytechnic Normal University,Guangzhou 510262,China

2.State Key Lab. of Software Engineering,Wuhan University,Wuhan 430072,China

Abstract

The disaccords and the redundancy features of a sample dataset will drop the classification quality and efficiency. In this paper,the method called consistent feature selection reduction is proposed about the classification data set.This method group together the inconsistent datum of the best possible category and make the data set uniform based on the Bayesian formula and a threshold value.Then a category distinguish matrix is built upon the consistent data set and the least feature variable subset that can distinguish the classification accurately is obtained through the category distinguish matrix.A heuristic search strategy and a practical example are given.The result shows the consistent feature selection reduction method can eliminate the disaccords of the sample dataset,select the optimal feature variables,drop the dimension of the data and reduce the redundancy information effectively.

Key words [data mining](#) [classification](#) [feature selection](#) [data reduction](#)

DOI:

通讯作者 吴新玲

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(721KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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

· [吴新玲](#)