

本期目录 | 下期目录 | 过刊浏览 | 高级检索

[打印本页] [关闭]

论文

应用特征项分布信息的信息增益改进方法研究

杨玉珍,刘培玉\*,朱振方,邱焯

山东师范大学信息科学与工程学院, 山东 济南 250014

摘要:

在特征项分布不平衡的情况下,传统信息增益算法的分类性能会急剧下降,针对此缺陷提出了一种利用特征项分布信息来改进信息增益公式的计算方法。通过计算特征项分布信息来判定特征项是否存在不平衡性,并利用此信息来平衡特征项不出现时对分类精度的影响。通过实验验证,改进后的计算方法整体上比传统的信息增益算法具有更好的性能。

关键词: 特征选择;信息增益;类内离散度;类间离散度

Research of an improved information gain method using distribution information of terms

杨玉珍,刘培玉\*,朱振方,邱焯

Department of Information Science and Engineering, Shandong Normal University, Jinan 250014, Shandong, China

Abstract:

Classification performance of a traditional information gain algorithm will rapidly decline when feature items are in an unbalanced distribution. An improved calculation method of an information gain formula using feature items' distribution information is proposed. Distribution information of feature items is computed to judge whether the imbalance of feature items exists and balance the influence of classification accuracy when the feature items do not appear. The improved calculation method has better performance through the experiment.

Keywords: feature selection; information gain; distribution Information inside a class; distribution Information among classes

收稿日期 2009-07-07 修回日期 网络版发布日期

DOI:

基金项目:

国家自然科学基金资助项目(60873247)

通讯作者: 刘培玉(1960 ),教授,博士生导师,主要研究方向计算机网络信息安全、网络系统规划、网络信息资源开发和软件开发技术。

作者简介:

本刊中的类似文章

扩展功能

本文信息

Supporting info

PDF(432KB)

[HTML全文]

[\({article.html\\_WenJianDaXiao} KB\)](#)

参考文献[PDF]

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

引用本文

Email Alert

本文关键词相关文章

特征选择;信息增益;类内离散度;类间离散度

本文作者相关文章

杨玉珍

刘培玉

朱振方

邱焯