

数据库、信息处理

随机森林针对小样本数据类权重设置

李建更, 高志坤

北京工业大学 人工智能与机器人研究所, 北京 100124

收稿日期 2008-5-15 修回日期 2008-9-1 网络版发布日期 2009-9-15 接受日期

摘要 随机森林已经被证明是一种高效的分类与特征选择方法。尽管参数的设置对结果影响较小, 但合适的参数可以使分类器得到理想的效果。主要针对癌症研究中小样本不均衡数据的分类和特征选择问题, 研究了随机森林中类权重的设置。为了比较在不同的类权重下特征选择的效果, 同时使用支持向量机 (Support Vector Machine, SVM) 方法。最终结果显示最优的类权重是不确定的。最后总结出几条规律指导研究者选择合适的权重使分类和特征选择效果得到改善。

关键词 [随机森林](#) [类权重](#) [小样本](#) [支持向量机](#) [特征选择](#)

分类号 [TP391](#)

Setting of class weights in random forest for small-sample data

LI Jian-geng, GAO Zhi-kun

Institute of Artificial Intelligence and Robotics, Beijing University of Technology, Beijing 100124, China

Abstract

Random forest has been proved to be an efficient algorithm for classification and feature selection in bioinformatics. Although the effect of parameter setting on results is very limited, a group of appropriate parameters can generate excellent performance. This paper focuses on the setting of class weights in random forest to deal with classification and feature selection problems of unbalanced small-sample data and determines the optimal class weight. In order to compare the performance of feature selection with different weights, SVM is applied in the paper. The results show that optimal class weight is variable and cannot form a standard. However, people can find some weights with which not only classification but also feature selection can get better performance.

Key words [random forest](#) [class weight](#) [small-sample](#) [Support Vector Machine \(SVM\)](#) [feature selection](#)

DOI: 10.3778/j.issn.1002-8331.2009.26.038

通讯作者 李建更 gaozhikun1314@emails.bjut.edu.cn

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(793KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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“随机森林”的相关文章](#)
- ▶ 本文作者相关文章
- [李建更](#)
- [高志坤](#)