

工程与应用

## Boosting算法在基因表达谱样本分类中的应用

刘全金<sup>1</sup>, 李颖新<sup>2</sup>

1. 安庆师范学院 物理与电气工程学院, 安徽 安庆246011

2. 北京经纬纺机新技术有限公司 CCD部, 北京 100176

收稿日期 2007-8-24 修回日期 2007-10-24 网络版发布日期 2008-5-6 接受日期

**摘要** 基于基因表达谱结构提出一种基因表达谱的样本分类方法。首先用基因的Bhattacharyya距离衡量其所含样本类别的信息, 过滤Bhattacharyya距离较小的噪声基因; 然后修改重复剪辑近邻算法, 剔除噪声样本; 再基于Boosting算法构建支持向量机组合分类器; 最后以结肠癌基因表达谱样本为例, 进行了分类实验。实验结果表明该方法简单、有效, 对基因表达谱样本的分类问题有强的实用性。

**关键词** [Bhattacharyya距离](#) [重复剪辑近邻法](#) [Boosting算法](#)

分类号

## Application of Boosting algorithm to sample categorization of gene expression profiles

LIU Quan-jin<sup>1</sup>, LI Ying-xin<sup>2</sup>

1. School of Physics & Electronic Engineering, Anqing Teachers College, Anqing, Anhui 246011, China

2. CCD Item, Beijing Jingwei Textile Machinery New Technology Co., LTD, Beijing 100176, China

### Abstract

In this paper an approach is proposed for sample categorization of gene expression profiles based on structure of gene expression profiles. Firstly, genes are removed as "noise genes" with small Bhattacharyya distance. Secondly, multi-edit-nearest-neighbor algorithm is modified to eliminate "noise samples". Then boosting-based support vector machines combination classifiers are constructed and employed to classify the samples. Finally, this methods is used to classify colon genes expression profiles samples. The results show that the means is feasible and effective.

**Key words** [Bhattacharyya distance](#) [multi-edit-nearest-neighbor algorithm](#) [Boosting algorithm](#)

DOI:

通讯作者 刘全金 [liuquanjing2002@yahoo.com.cn](mailto:liuquanjing2002@yahoo.com.cn)

### 扩展功能

#### 本文信息

▶ [Supporting info](#)

▶ [PDF\(527KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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 "Bhattacharyya距离" 的相关文章](#)

▶ 本文作者相关文章

· [刘全金](#)

· [李颖新](#)