

博士论坛

## 基于基因表达式编程抽取特征的分类算法

姜 玥<sup>1,2</sup>, 唐常杰<sup>1</sup>, 吴 江<sup>1</sup>, 叶尚玉<sup>1</sup>, 陈 瑜<sup>1</sup>

1. 四川大学 计算机学院, 成都 610064

2. 西南民族大学 计算机科学与技术学院, 成都 610041

收稿日期 修回日期 网络版发布日期 2007-8-29 接受日期

**摘要** 传统的基因表达式编程在解决多分类问题时,人为地把多分类转换成多次两分类。融合了K-最邻近分类和基因表达式编程两种技术,做了下列工作:(1)提出了基于基因表达式编程中的特征抽取,证明了特征抽取区多样性定理;(2)提出了特征的自动聚类策略和特征集的自动选择策略,用特征的聚类辅助对象的分类;(3)提出基于基因表达式编程的最邻近距离分类算法,用抽取出的特征采用最邻近距离分类算法进行多分类;(4)实验表明,采用基于基因表达式编程的最邻近距离分类算法,有效地解决了多分类问题,改善了分类性能,使平均分类正确率提高约4%~10%,用于分类的特征维数减少60%~79%。

**关键词** [基因表达式编程](#) [多分类问题](#) [特征抽取](#)

分类号

## New method for classification: extract features based on gene expression programming

JIANG Yue<sup>1,2</sup>, TANG Chang-jie<sup>1</sup>, WU Jiang<sup>1</sup>, YE Shang-yu<sup>1</sup>, CHEN Yu<sup>1</sup>

1. College of Computer Science, Sichuan University, Chengdu 610064, China

2. College of Computer Science & Technology, Southwest University for Nationalities, Chengdu 610041, China

### Abstract

Traditional gene expression programming method solves multi-classification by doing two-class classification in multi times. This paper fuses K-nearest neighbor classification algorithm and gene expression programming. The contributions of this paper include: (1) proposes the concepts of feature extraction based on gene expression programming, and proves the theorem of variety of feature extraction areas, (2) proposes automatic feature clustering strategy and automatic feature selection strategy to classify features using assistant object of clustering, (3) proposes K-nearest neighbor classification algorithm based on gene expression programming. It works successfully in multi-classification by applying KNN on features, (4) demonstrates effectiveness of new proposed GEP-KNN algorithm by extensive experiments in multi-classification task. The average classification accuracy is increased by about 4%~10% and the feature dimensions of classification is decreased by 60%~79%.

**Key words** [gene expression programming](#) [multi-classification problem](#) [feature extraction](#)

DOI:

通讯作者 姜 玥 [E-mail: jiangyue@cs.scu.edu.cn](mailto:jiangyue@cs.scu.edu.cn)

### 扩展功能

#### 本文信息

▶ [Supporting info](#)

▶ [PDF\(1950KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“基因表达式编程” 的相关文章](#)

▶ [本文作者相关文章](#)

· [姜 玥](#)

· [唐常杰](#)

· [吴 江](#)

· [叶尚玉](#)

· [陈 瑜](#)