

论文

基于Boosting算法集成遗传模糊分类器的文本分类

罗军 况奔

重庆教育学院 重庆教育学院

摘要:

提出一种新颖的基于Boosting模糊分类的文本分类方法。首先采用潜在语义索引(LSI)对文本特征进行选择;然后提出Boosting算法集成模糊分类器学习,在每轮迭代训练过程中,算法通过调整训练样本的分布,利用遗传算法产生分类规则。减少分类规则能够正确分类样本的权值,使得新产生的分类规则重点考虑难于分类的样本。实验结果表明,该文本分类算法具有良好分类的性能。

关键词: 模糊分类 特征选择 潜在语义索引 Boosting算法 文本分类

Text categorization based on genetic fuzzy classification and Boosting method

Abstract:

A novel method for text categorization, which is based on boosting fuzzy classification, was proposed in the paper. Latent Semantic Index (LSI) was used to select text feature and then Boosting algorithm was proposed to integrate fuzzy classification. In each iteration training of boosting algorithm, the distribution of training instances was adjusted, and classification rules were created by genetic algorithm. The weights of the training instances that were classified correctly by available rules were reduced, so that the new fuzzy rule focuses on the misestimate or uncovered instances. Experimental results show that classifier based on fuzzy classification is effective and efficient.

Keywords: fuzzy classification feature selection Latent Semantic Index (LSI) Boosting algorithm text categorization

收稿日期 2008-03-25 修回日期 2008-05-09 网络版发布日期

DOI:

基金项目:

通讯作者: 罗军

作者简介:

参考文献:

本刊中的类似文章

- 1. 田学东 郝楠.基于模糊分类的印刷体数学公式抽取方法[J]. 计算机应用, 2007,27(8): 2036-2037

文章评论 (请注意:本站实行文责自负, 请不要发表与学术无关的内容!评论内容不代表本站观点.)

扩展功能

本文信息

- Supporting info
PDF(768KB)
[HTML全文]
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
文章反馈
浏览反馈信息

本文关键词相关文章

- 模糊分类
特征选择
潜在语义索引
Boosting算法
文本分类

本文作者相关文章

- 罗军
况奔

PubMed

- Article by
Article by

Feedback form with fields for name and email address.

反馈  
标题

验证码

1468