

数据库、信号与信息处理

贝叶斯文本分类器的研究与改进

史瑞芳

山西经济管理干部学院, 太原 030024

收稿日期 2008-9-25 修回日期 2009-1-8 网络版发布日期 2009-4-20 接受日期

摘要 朴素贝叶斯文本分类是目前公认的一种简单有效的概率分类方法, 但该方法的数据稀疏问题以及所采用的 Laplace 平滑方法还不是最优, 存在一定的缺陷。因此, 用一元统计语言模型的平滑方法来改进数据稀疏状况, 提高了分类效果。

关键词 [贝叶斯文本分类](#) [数据稀疏](#) [平滑](#)

分类号

Research and improvement on Naive Bayes text classifier

SHI Rui-fang

Shanxi Economic Management Institute, Taiyuan 030024, China

Abstract

Naive Bayes method is a simple and effective established probability categorization method at present. However, the problems on scattered data in methodology and Laplace smoothness method have some disadvantages. Therefore, author proposes to use uni-gram smoothness method to improve the condition and the effect on categorization by Bayes method.

Key words [Bayes text categorization](#) [scattered data](#) [smoothness](#)

DOI: 10.3778/j.issn.1002-8331.2009.12.048

通讯作者 史瑞芳 srfang0620@163.com

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(412KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“贝叶斯文本分类”的相关文章](#)

▶ [本文作者相关文章](#)

· [史瑞芳](#)