

数据库、信号与信息处理

## 改进的朴素贝叶斯垃圾邮件过滤算法

翟军昌<sup>1,2</sup>, 秦玉平<sup>2</sup>, 王春立<sup>3</sup>

- 1.渤海大学 公共计算机教研部, 辽宁 锦州 121000
- 2.渤海大学 信息科学与工程学院, 辽宁 锦州 121000
- 3.大连海事大学 计算机科学与技术学院, 辽宁 大连 116023

收稿日期 2008-11-3 修回日期 2009-1-15 网络版发布日期 2009-5-8 接受日期

**摘要** 介绍了朴素贝叶斯垃圾邮件过滤算法, 对于朴素贝叶斯算法中条件概率的计算, 选用了多变量贝努里事件模型的计算方法, 在多变量贝努里事件模型的基础上进行了改进, 并在Ling-Spam语料库上进行实验, 实验结果表明改进后的算法有效地提高了过滤器的召回率和精确率, 并且降低了过滤器的错误率。

**关键词** [垃圾邮件](#) [朴素贝叶斯](#) [信息增益](#) [多变量贝努里事件模型](#)

分类号

## Improved Naïve Bayesian spam filtering algorithm

ZHAI Jun-chang<sup>1,2</sup>, QIN Yu-ping<sup>2</sup>, WANG Chun-li<sup>3</sup>

- 1.Dept. of Public Computer Teaching & Research, Bohai University, Jinzhou, Liaoning 121000, China
- 2.College of Information Science and Technology, Bohai University, Jinzhou, Liaoning 121000, China
- 3.School of Computer Science and Technology, Dalian Maritime University, Dalian, Liaoning 116023, China

### Abstract

The paper describes the Naïve Bayesian spam filtering algorithms. In terms of probability calculation of Naïve Bayes algorithm, the paper selects calculation of multi-variable model of Bernoulli event, and makes improvements to multi-variable model of Bernoulli event, and carries out an experimental on the Ling-Spam corpus. The results show that the improved algorithm can effectively enhance the recall and accuracy of the filter and lower the error rate of the filter.

**Key words** [spam email](#) [Naï ve Bayes](#) [information gain](#) [multi-variable model of Bernoulli events](#)

DOI: 10.3778/j.issn.1002-8331.2009.14.044

通讯作者 翟军昌 [zhaijunchang@163.com](mailto:zhaijunchang@163.com)

### 扩展功能

#### 本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(683KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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“垃圾邮件”的相关文章](#)
- ▶ 本文作者相关文章

- [翟军昌](#)
- [秦玉平](#)
- [王春立](#)