

数据库、信号与信息处理

基于贝叶斯方法和信息指纹的博客评论过滤

马如林, 蒋华, 张庆霞

桂林电子科技大学 计算机与控制学院, 广西 桂林 541004

收稿日期 2007-10-26 修回日期 2008-1-16 网络版发布日期 2008-8-19 接受日期

摘要 博客的出现丰富和改变了网络的内涵, 影响了人们的信息传递方式, 同时博客评论作为一种交互方式在博客中广泛存在, 给信息监管带来了新的问题。通过分析现有的博客过滤系统, 将广泛应用于文本过滤的贝叶斯方法应用到博客评论中, 针对博客评论中广泛存在的广告机器人特点, 结合信息指纹对其进行识别和过滤。同时对影响博客评论过滤效果和执行速度的指纹函数进行了分析讨论和实验对比, 实验结果表明基于贝叶斯方法和信息指纹相结合的博客评论过滤是行之有效的, 而且相对于单独的贝叶斯方法更有利于提高系统运行效率和发现广告机器人现象。

关键词 [博客](#) [贝叶斯](#) [评论](#) [信息指纹](#)

分类号

Blog's content filtering based on Bayes method and information fingerprint

MA Ru-lin,JIANG Hua,ZHANG Qing-xia

School of Computer and Control, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

Abstract

The appearance of blog enriches and changes the network's connotation, and influences the ways of information-delivering. Blog criticism, as an exchanging way, has been widely used in blog and thus brings new problems to information warding. This paper on one hand, applies Bayes of text filtering in blog criticism by analysis of blog filtering system in hand; On the other hand, because of the specific features of robot widely existing in blog criticism, this paper recognizes and filters the criticism combining the information fingerprint. Moreover, this paper analyzes and discusses the fingerprint functions that influence blog-filtering's effect and carrying-out speed. The result of this experiment shows that this blog-filtering is effective, based on Bayes and information fingerprint, and is more advanced than the only Bayes in improving system running efficiency and finding out the phenomenon of advertisement robot.

Key words [blog](#) [Bayes](#) [comments](#) [information fingerprint](#)

DOI: 10.3778/j.issn.1002-8331.2008.24.048

通讯作者 马如林 rulinma@gmail.com

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(646KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“博客”的 相关文章](#)

▶ 本文作者相关文章

· [马如林](#)

· [蒋华](#)

· [张庆霞](#)