

数据库、信息处理

词间相关性在贝叶斯文本分类中的应用研究

章舜仲^{1,2}, 王树梅¹, 黄河燕³, 陈肇雄³

1.南京理工大学 计算机科学系, 南京 210094

2.南京财经大学 电子商务系, 南京 210046

3.中国科学院 计算机语言信息工程研究中心, 北京 100083

收稿日期 2008-4-1 修回日期 2008-6-11 网络版发布日期 2009-5-27 接受日期

摘要 针对朴素贝叶斯分类的属性独立性假设的不足, 讨论了相关性及多变量相关的概念, 给出词间相关度的定义。在TAN分类器的词间相关性分析基础上, 提出一种文档特征词相关度估计公式及其在改进朴素贝叶斯分类模型中应用的算法, 在Reuters-21578文本数据集上的实验表明, 改进算法简单易行, 能有效改进贝叶斯分类性能。

关键词 [文本分类](#) [朴素贝叶斯](#) [事件相关](#) [相关度](#) [树扩展型朴素贝叶斯分类器](#)

分类号

Research on application of word correlation in Naive Bayes text classification

ZHANG Shun-zhong^{1,2}, WANG Shu-mei¹, HUANG He-yan³, CHEN Zhao-xiong³

1.Department of Computer Science, Nanjing University of Science and Techology, Nanjing 210094, China

2.Department of Electronic Business, Nanjing University of Finance and Economics, Nanjing 210046, China

3.Computer Language Information Engineering Research Center, Chinese Academy of Sciences, Beijing 100083, China

Abstract

Aiming at the deficiency of Naive Bayes' attribute independence assumption, the concept of correlation and that between multi-variations were discussed, and the definition of correlation degree between terms was presented. Based on the analysis of the correlation between terms of TAN classifier, authors proposed a formula to evaluate the correlation degree between document feature words and the algorithm of its application to ameliorating Naive Bayesian classifier. The experiments on Reuters-21578 collection show the improvement of algorithm to be simple, effective and easy to implement.

Key words [text classification](#) [Naive Bayes](#) [event correlation](#) [correlation degree](#) [Tree Augmented Naive Bayes \(TAN\) classifier](#)

DOI: 10.3778/j.issn.1002-8331.2009.16.046

通讯作者 章舜仲 zszs@jlonline.com

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(527KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“文本分类”的相关文章](#)

▶ 本文作者相关文章

· [章舜仲](#)

· [王树梅](#)

· [黄河燕](#)

· [陈肇雄](#)