

数据库、信号与信息处理

基于词对向量空间模型的新事件检测方法

樊旭琴¹, 张永奎^{1, 2}

1.山西大学 计算机与信息技术学院, 太原 030006

2.山西大学 计算智能与中文信息处理省部共建教育部重点实验室, 太原 030006

收稿日期 2009-4-8 修回日期 2009-6-8 网络版发布日期 2010-4-21 接受日期

摘要 新事件检测(NED)的目标是从一个或多个新闻源中检测出报道一个新闻话题的第一个新闻。传统向量空间模型采用单个词来表示文本特征,考虑到词的位置信息以及其他的表示内容的信息,提出了词对表示文本的方法,并结合HowNet资源对所抽取的词对进行归一化处理,最后对不同类别新闻中不同词性对的权重参数进行优化。通过在已有的突发性新闻语料上进行实验,表明这种改进方法的效果比较明显,性能也有一定的提高。

关键词 [向量空间模型](#) [词对特征](#) [新事件检测](#)

分类号 [TP391](#)

New event detection method based on word pairs vector space model

FAN Xu-qin¹, ZHANG Yong-kui^{1, 2}

1.School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

2.Key Laboratory of MOE for Computation Intelligence and Chinese Information Processing, Shanxi University, Taiyuan 030006, China

Abstract

New Event Detection (NED) aims at detecting the first news item on one topic from one or more news reports. The traditional vector space model adopts single word to represent the text features, considering the information of word position and other information of expressing content, this paper proposes an approach using word pairs to express text content. Combined with the HowNet, the extracted word pairs are normalized. Then the different weight parameters of different part of speech pairs are given according to different types of news reports. Experiments on emergency news corpus show that the word-pair method can significantly improve the representation results.

Key words [vector space model](#) [word pair feature](#) [new event detection](#)

DOI: 10.3778/j.issn.1002-8331.2010.12.036

通讯作者 樊旭琴 fxq514@126.com

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(604KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“向量空间模型”的相关文章](#)

▶ [本文作者相关文章](#)

· [樊旭琴](#)

· [张永奎](#)

·