

周学广,高飞,孙艳.基于依存连接权VSM的子话题检测与跟踪方法[J].通信学报,2013,(8):1~9

基于依存连接权VSM的子话题检测与跟踪方法

Sub-topic detection and tracking based on dependency connection weights for vector space model

投稿时间: 2012-05-15

DOI: 10.3969/j.issn.1000-436x.2013.08.001

中文关键词: [话题检测与跟踪](#) [依存连接权](#) [关联词对](#) [报道关系检测](#) [向量空间模型](#)

英文关键词: [topic detection and tracking](#) [dependency connection weights](#) [associating words group](#) [report relation detection](#) [vector space model](#)

基金项目:海军工程大学科学研究基金资助项目(HGDYDJ10008)

作者

[周学广](#), [高飞](#), [孙艳](#)

单位

[海军工程大学 信息安全系, 湖北 武汉430033](#)

摘要点击次数: 503

全文下载次数: 477

中文摘要:

针对在新闻话题中报道突发、热点相似且子话题层次丰富的现象,依据增量TF-IDF值构造特征维,生成全局向量;然后在时间窗内生成特征连接权的局部邻接图,利用依存句法进行分析降维;最后采用领域词典加权,时间阈值衰减;从而构造出利用依存连接权VSM进行关联分析的子话题检测与跟踪(sTDT)计算方法。实验表明,利用依存关联分析使文本表示由线性变为平面结构,能够有效地提取描述子话题;在人工标注的测试语料下,其最小DET代价比经典方法至少降低2.2%。

英文摘要:

Aiming at the phenomenon that there are abrupt reports, similar topics and abundant levels of subtopics in the news, a novel method based on relationship analysis using dependent sentence pattern was proposed for sub-topic detection and tracking (sTDT), which constructed feature dimensions to generate the global vectors according to the increment of TF-IDF, and then created the partial adjoint map based on the connection weights within the time window and decreased the dimensions through dependent sentence pattern. Finally, a novel method for sTDT computing was built with adjoint dictionary weights and time threshold attenuation. Experiments show that the proposed method transfers the text from linear to plane structure, and extracts the subtopics effectively, of which the minimum DET cost is reduced by at least 2.2 percent than that of classical methods.

[查看全文](#) [查看/发表评论](#) [下载PDF阅读器](#)

关闭

版权所有:《通信学报》

地址:北京市丰台区成寿寺路11号邮电出版大厦8层 电话:010-81055478, 81055479

81055480, 81055482 电子邮件: xuebao@ptpress.com.cn

技术支持:北京勤云科技发展有限公司