

数据库、信号与信息处理

## 基于SVM的中文报道关系识别方法研究

王强<sup>1</sup>, 张永奎<sup>2</sup>

1.山西大学 计算机与信息技术学院, 太原 030006

2.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 2008-7-2 修回日期 2008-9-26 网络版发布日期 2008-11-18 接受日期

**摘要** 针对网络新闻的特点, 从人名、时间名、地点名、组织机构名、内容五个方面抽取特征词形成特征向量。在此基础上, 分别进行了相似度计算, 其中, 人名、组织机构名、内容采用余弦夹角的方法, 时间和地点向量, 相似度计算采用了引入报道时间和关联度计算。最后, 使用这5个相似度作为特征, 使用SVM进行训练, 并在测试集上进行了测试。测试结果表明, 这种方法可以有效地改善系统的性能。

**关键词** [报道关系识别](#) [话题检测与跟踪](#) [多向量表示模型](#)

分类号

## Research on Chinese story link detection based on SVM

WANG Qiang<sup>1</sup>, ZHANG Yong-kui<sup>2</sup>

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

2.Key Lab of Ministry of Education for Computation Intelligence and Chinese Information Processing, Taiyuan 030006, China

### Abstract

Via analyzing the characteristic of news in the Web, construct the feature vector using features from five entity categories: persons, time, location, organizations, and content. Using story time and entity relatedness for temporal or place vector when calculating their similarity and cosine similarity for others. All the features together with the entity relatedness are integrated by Support Vector Machine (SVM). Experimental results show that this method can improve system performance effectively.

**Key words** [story link detection](#) [topic detection and tracking](#) [multi-vector mode](#)

DOI: 10.3778/j.issn.1002-8331.2008.33.044

通讯作者 王强 [wangqiang25719@163.com](mailto:wangqiang25719@163.com)

### 扩展功能

#### 本文信息

▶ [Supporting info](#)

▶ [PDF\(572KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“报道关系识别”的  
相关文章](#)

▶ 本文作者相关文章

· [王强](#)

· [张永奎](#)