

数据库、信号与信息处理

中文文本分类中利用依存关系的实验研究

王 鹏, 樊兴华

重庆邮电大学 计算机科学与技术研究所, 重庆 400065

收稿日期 2008-8-5 修回日期 2008-10-28 网络版发布日期 2010-1-28 接受日期

摘要 为了利用依存关系进行短文本分类, 研究了利用依存关系进行短文本分类存在的四个关键问题。分别在长文本语料集和两个短文本语料集上, 抽取具有依存关系的词对, 并利用这些词对作为特征进行分类实验。实验结果表明: 依存关系能够作为有效的特征进行文本分类, 并能够改善文本分类的性能; 单独把依存关系作为特征, 不能提高短文本的分类性能; 可以利用依存关系作为特征扩充的手段, 增加短文本的特征, 增强短文本的描述能力, 进而进行有效的短文本分类。

关键词 [依存关系](#) [短文本](#) [文本分类](#)

分类号 [TP18](#)

Study on Chinese text classification based on dependency relation

WANG Peng, FAN Xing-hua

Institute of Computer Science and Technology, Chongqing University of Post and Telecommunication, Chongqing 400065, China

Abstract

Four key issues of classifying Chinese short text using dependency relation are discussed to use dependency relation to classify Chinese short text effectively. This paper extracts the dependency relations between two words in a long-text corpus and two short-text corpuses, and uses these word-pairs to classify texts in order to analyze the role of dependency relation in short text classification. Experiments show that Using dependency relation to classify texts can improve the classification performance; Using dependency relation to classify short texts lonely can not improve the classification performance; dependency relation as means to expand features can increase features and enhance description ability of short text in order to classify short texts effectively.

Key words [dependency relation](#) [short text](#) [text classification](#)

DOI: 10.3778/j.issn.1002-8331.2010.03.039

通讯作者 王 鹏 hmilywp26@gmail.com

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(571KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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“依存关系”的相关文章](#)
- ▶ 本文作者相关文章
 - [王 鹏](#)
 - [樊兴华](#)