

数据库、信息处理

基于String Kernel和KPCA的负实例语法特征提取算法

吕威^{1, 2}, 林文昶¹, 姚正安¹, 李磊¹

1.中山大学 软件研究所, 广州 510275

2.北京师范大学 珠海分校 信息技术学院, 广东 珠海 519085

收稿日期 2008-10-10 修回日期 2008-11-18 网络版发布日期 2009-7-9 接受日期

摘要 提出通过String Kernel方法把负实例语法数据库中的负实例转化成核矩阵, 再用Kernel Principal Component Analysis (KPCA) 对转换的核矩阵进行特征提取, 进而可将原始负实例数据库按照这些特征分成多个容量较小的特征表。通过构造负实例特征索引表设计了一个分类器, 待检查的句子通过此分类器被分配到某个负实例特征表里进行匹配搜索, 而此特征表的特征属性数和记录数要远远小于原始负实例数据库中的相应数目, 从而大大提高了检查的速度, 同时不影响语法检查的精度。通过比较测试, 可看出提出的方法在保证语法检查精度的同时有更快的速度。

关键词 [String Kernel](#) [核主成分分析](#) [负实例](#) [特征提取](#)

分类号

False instance grammatical feature extraction algorithm based on String Kernel and KPCA

LV Wei^{1, 2}, LIN Wen-chang¹, YAO Zheng-an¹, LI Lei¹

1.Software Research Institute, Zhongshan University, Guangzhou 510275, China

2.School of Information Technology, Beijing Normal University Zhuhai Campus, Zhuhai, Guangdong 519085, China

Abstract

This paper presents a method that translates false instance in grammatical database to kernel matrix through String Kernel, and uses KPCA to extract feature of the translated kernel matrix. We can separate the original false instance database into many small characteristic tables according to these extracted features, and design a classified machine by constructing false instance characteristic table. A new sentence is distributed to some characteristic table for matching of false instance through this classification machine. For characteristic table is much little than original false instance database, the running speed is enhanced very much without decreasing the accuracy of grammatical check. By compared with grammar inspection function of word, the new system demonstrates more quick speed while keeping the accuracy of grammatical check

Key words [String Kernel](#) [Kernel Principal Component Analysis \(KPCA\)](#) [false instance](#) [feature extraction](#)

DOI: 10.3778/j.issn.1002-8331.2009.20.041

通讯作者 luwei00@126.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“String Kernel” 的相关文章](#)

▶ [本文作者相关文章](#)

· [吕威](#)

· [林文昶](#)

· [姚正安](#)

· [李磊](#)