

机器学习

## 基于SVM-Adaboost的中文组块分析

别致<sup>1</sup>,周俊生<sup>2</sup>,陈家骏<sup>1</sup>

1.南京大学 计算机软件新技术国家重点实验室, 南京 210093

2.南京师范大学 计算机科学系, 南京 210097

收稿日期 2008-4-30 修回日期 2008-5-29 网络版发布日期 2008-7-17 接受日期

**摘要** 组块分析是一种非常重要的句法分析预处理手段, 通过将文本划分成一组互不重叠的片断, 来达到降低句法分析的难度。提出一种基于SVM-Adaboost的中文组块分析方法, 将基于线性核函数的支持向量机与Adaboost算法相结合, 以基于线性核函数的SVM作为Adaboost的分量分类器, 在学习过程中改变分量分类器的核参数。实验结果表明了该算法的有效性。

**关键词** [中文组块分析](#) [Adaboost](#) [支持向量机](#)

分类号

## SVM-Adaboost based Chinese text chunking

BIE Zhi<sup>1</sup>, ZHOU Jun-sheng<sup>2</sup>, CHEN Jia-jun<sup>1</sup>

1.State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China

2.Department of Computer, Nanjing Normal University, Nanjing 210097, China

### Abstract

Text chunking is a very important approach to preprocessing parsing. It divides text into syntactically related non-overlapping groups of chunks in order to reduce the complexity of the full parsing. In this paper, a SVM-Adaboost algorithm is applied for Chinese text chunking which combines Adaboost with linear-kernel SVM. This algorithm uses SVM as weak learners for AdaBoost and adjusts the kernel parameter of SVM in the learning process. The experimental results show that it is an effective approach.

**Key words** [Chinese text chunking](#) [Adaboost](#) [support vector machine](#)

DOI: 10.3778/j.issn.1002-8331.2008.21.047

通讯作者 别致 [biez@nlp.nju.edu.cn](mailto:biez@nlp.nju.edu.cn)

### 扩展功能

#### 本文信息

▶ [Supporting info](#)

▶ [PDF\(576KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“中文组块分析” 的相关文章](#)

▶ 本文作者相关文章

· [别致](#)

· [周俊生](#)

· [陈家骏](#)