

数据库与信息处理

基于分类回归树CART的汉语韵律短语边界识别

钱揖丽^{1,2}, 荀恩东³

1.北京工业大学 计算机学院, 北京 100022

2.山西大学 计算机与信息技术学院, 太原 030006

3.北京语言大学 信息科学学院, 北京 100083

收稿日期 2007-11-8 修回日期 2007-12-25 网络版发布日期 2008-2-11 接受日期

摘要 提出了一种基于分类回归树(Classification And Regression Tree, CART)的汉语韵律短语识别方法。该方法从语音流中提取与韵律短语边界有关的声学特征,从文本中提取短语边界的语言学特征,并将两类特征有机结合构成CART特征集,建立CART决策模型。开放测试结果显示,利用该CART模型在词边界中识别韵律短语边界,其识别准确率平均可达95.91%。

关键词 [分类回归树](#) [决策树](#) [韵律短语](#) [边界](#)

分类号

Identification of Chinese prosodic phrase based on CART

QIAN Yi-li^{1,2}, XUN En-dong³

1.College of Computer Science and Technology, Beijing University of Technology, Beijing 100022, China

2.College of Computer Science and Information Technology, Shanxi University, Taiyuan 030006, China

3.College of Information Sciences, Beijing Language and Culture University, Beijing 100083, China

Abstract

This paper presents a CART-based method for identifying the Chinese prosodic phrase. Firstly, it obtains acoustic characteristics which have relation to the boundary of prosodic phrase from speech, and it gain linguistic characteristics of prosodic phrase boundary from text. Secondly, it combines these characteristics effectively to construct characteristic muster, and then use it to build CART model. The results of opening test show that identifying the boundary of Chinese prosodic phrase using this CART model, its precision can reach 95.91% averagely.

Key words [CART](#) [decision tree](#) [prosodic phrase](#) [boundary](#)

DOI:

通讯作者 钱揖丽 qianyl@emails.bjut.edu.cn

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(599KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“分类回归树” 的相关文章](#)

▶ 本文作者相关文章

· [钱揖丽](#)

· [荀恩东](#)