

数据库、信息处理

基于最大熵模型的语义块切分

谢法奎^{1,2}, 张全²

1.中国科学院 研究生院, 北京 100039

2.中国科学院 声学研究所, 北京 100190

收稿日期 2008-5-19 修回日期 2008-8-12 网络版发布日期 2009-9-15 接受日期

摘要 语义块切分是HNC理论的重要课题, 与以往的处理策略不同, 采用统计建模的方法来解决这一问题。采用词语、词性、概念等信息组成特征模板, 并应用增量方法进行特征选择, 构建了一个基于最大熵模型的语义块切分系统。在HNC标注语料库上的测试取得了较好的效果, 开放测试的正确率和召回率分别达到了83.78%和91.17%。

关键词 [最大熵模型](#) [语义块](#) [概念层次网络](#)

分类号 [TP391](#)

Semantic chunks segmentation based on maximum entropy model

XIE Fa-kui^{1,2}, ZHANG Quan²

1.Graduate University of Chinese Academy of Sciences, Beijing 100039, China

2.Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China

Abstract

Semantic chunks segmentation is an important task in the Hierarchical Network of Concepts (HNC) theory. To deal with this problem, this paper adopts a new method based on statistical modeling. And forms some feature templates with word, POS, concept, and selects features by an incremental way. Finally, construct a semantic chunks segmentation system based on a maximum entropy model. The experiment is taken on HNC corpus, and the result shows that the model works well, the open test precision and recall are 83.78% and 91.17% respectively.

Key words [maximum entropy model](#) [semantic chunk](#) [Hierarchical Network Concepts \(HNC\)](#)

DOI: 10.3778/j.issn.1002-8331.2009.26.035

通讯作者 谢法奎 holinax@tom.com

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(610KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“最大熵模型”的相关文章](#)

▶ [本文作者相关文章](#)

· [谢法奎](#)

·

· [张全](#)