

论文

利用覆盖歧义检测法和统计语言模型进行汉语自动分词

王显芳,杜利民

中国科学院声学研究所语音交互信息技术研究中心,北京,100080

收稿日期 2002-3-19 修回日期 2002-8-9 网络版发布日期 2008-6-25 接受日期

摘要

该文探讨了利用覆盖歧义检测法和统计语言模型进行汉语自动分词的问题。采用了多次迭代的方法来进行汉语词层面统计语言模型的训练。该方法能够得到更优化的语言模型。该文详细介绍了统计语言模型的训练过程,给出了语言模型复杂度随迭代次数增加而减小的实验结果。还给出了在不同的统计语言模型阶数下切分正确率变化的情况,分析了切分正确率变化的原因。

关键词 [统计语言模型](#) [覆盖歧义检测法](#) [自动分词](#)

分类号 [TP391.4](#) [TN912.3](#)

Automatic Segmentation of Chinese using overlaying ambiguity examining method and statistics language model

Wang Xianfang, Du Limin

Center for Speech Interactive Information Technology Institute of Acoustics Chinese Academy of Sciences Beijing 100080 China

Abstract

In this paper, the question of Chinese automatic segmentation is discussed using overlaying ambiguity examining method and statistics language model. The multi-time iterative method is applied to train language model, which can produce a better model. The process of training language model is described in detail. The result shows that the perplexity of language model is reduced. The accuracy of segmentation changes with different language model and the reason is analyzed.

Key words [Statistics language model](#) [Overlaying ambiguity examining method](#) [Automatic segmentation](#)

DOI :

通讯作者

作者个人主页 王显芳;杜利民

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF \(538KB\)](#)

▶ [\[HTML全文\]\(OKB\)](#)

▶ [参考文献\[PDF\]](#)

▶ [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“统计语言模型”的相关文章](#)

▶ 本文作者相关文章

- [王显芳](#)
- [杜利民](#)