

人工智能及识别技术

基于前后文n-gram模型的古汉语句子切分

陈天莹<sup>1</sup>, 陈 蓉<sup>1</sup>, 潘璐璐<sup>1</sup>, 李红军<sup>1,2</sup>, 于中华<sup>1</sup>

(1. 四川大学计算机学院, 成都610064; 2. 西南科技大学计算机学院, 绵阳 621002)

收稿日期 修回日期 网络版发布日期 2007-1-29 接受日期

**摘要** 提出了基于前后文n-gram模型的古汉语句子切分算法, 该算法能够在数据稀疏的情况下, 通过收集上下文信息, 对切分位置进行比较准确的预测, 从而较好地处理小规模训练语料的情况, 降低数据稀疏对切分准确率的影响。采用《论语》对所提出的算法进行了句子切分实验, 达到了81%的召回率和52%的准确率。

**关键词** [n-gram模型](#) [数据稀疏](#) [平滑技术](#) [基于前后文的n-gram模型](#)

分类号

**DOI:**

通讯作者:

作者个人主页: [陈天莹<sup>1</sup>](#); [陈 蓉<sup>1</sup>](#); [潘璐璐<sup>1</sup>](#); [李红军<sup>1</sup>](#); [于中华<sup>1</sup>](#)

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(118KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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](#)
- ▶ [引用本文](#)
- ▶ [Email Alert](#)
- ▶ [文章反馈](#)
- ▶ [浏览反馈信息](#)

相关信息

- ▶ [本刊中 包含“n-gram模型”的相关文章](#)
- ▶ 本文作者相关文章
  - [陈天莹<sup>1</sup>, 陈 蓉<sup>1</sup>, 潘璐璐<sup>1</sup>, 李红军<sup>1,2</sup>, 于中华<sup>1</sup>](#)