

博士论文

基于CRF的百科全书文本段落划分

许 勇<sup>1</sup> , 宋 柔<sup>2</sup>

(1. 北京工业大学计算机科学学院, 北京 100022; 2. 北京语言大学计算机科学系, 北京 100083)

收稿日期 修回日期 网络版发布日期 2007-5-16 接受日期

**摘要** CRF模型是标注、切分序列数据的较新的概率模型, 在信息抽取等文本处理领域广受关注。该文介绍了CRF方法, 并将其应用到百科全书文本段落的划分上, 利用CRF的特征表述机制加入了文本单元序列中的长距离约束, 取得了比传统的隐马尔科夫方法更好的结果。

**关键词** [文本段落划分](#) [条件随机域模型](#) [隐马尔科夫模型](#)

**分类号** [TP18](#)

**DOI:**

通讯作者:

作者个人主页: [许 勇<sup>1</sup>](#) ; [宋 柔<sup>2</sup>](#)

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(193KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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“文本段落划分”的相关文章](#)
- ▶ 本文作者相关文章
  - [许 勇<sup>1</sup> , 宋 柔<sup>2</sup>](#)