

人工智能

基于多模板隐马尔可夫模型的文本信息抽取算法

胡宇舟¹;王雷¹;顾学道²

深圳现代计算机公司¹

收稿日期 2007-9-27 修回日期 网络版发布日期 2008-3-1 接受日期

摘要 由于训练数据来源的多样化,难以通过学习得到最优的模型参数,因此提出了一种基于多模板隐马尔可夫模型的文本信息抽取算法。该算法首先利用文本排版格式和分隔符等信息,对文本进行分块;然后在分块的基础上,对训练数据进行聚类以形成多个形式的模板(多模板),并对多模板数据训练得到隐马尔可夫初始概率及转移概率参数;最后,用被训练的数据统一训练释放概率参数,结合初始概率、转移概率以及释放概率参数对文本信息进行抽取。实验结果表明,该算法在精确度和召回率指标上比简单隐马尔可夫模型具有更好的性能。

关键词 [文本信息抽取](#) [隐马尔可夫模型](#) [多模板](#) [文本分块](#)

分类号

DOI:

对应的英文版文章: [A7095340](#)

通讯作者:

顾学道 guxd@mcm.com.cn

作者个人主页: [胡宇舟](#) [王雷](#) [顾学道](#)

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF \(806KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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“文本信息抽取”的相关文章](#)
- ▶ [本文作者相关文章](#)

- [胡宇舟](#)
- [王雷](#)
- [顾学道](#)