

人工智能及识别技术

基于种子自扩展的命名实体关系抽取方法

何婷婷, 徐超, 李晶, 赵君喆

(华中师范大学计算机科学与技术系, 武汉 430079)

收稿日期 修回日期 网络版发布日期 2006-10-27 接受日期

摘要 命名实体间关系的抽取是信息抽取中的一个重要研究问题, 该文提出了一种从大量的文本集合中自动抽取命名实体间关系的方法, 找出了所有出现在同一句子内、词语之间的距离在一定范围之内的命名实体对, 把它们的上下文转化成向量。手工选取少量具有抽取关系的命名实体对, 把它们作为初始关系的种子集合, 通过自学习, 关系种子集合不断扩展。通过计算命名实体对和关系种子之间的上下文相似度来得到所要抽取的命名实体对。通过扩展关系种子集合的方法, 抽取的召回率和准确率都得到了提高。该方法在对《人民日报》语料库的测试中, 取得了加权平均值F-Score为0.813的效果。

关键词 [命名实体](#) [关系抽取](#) [自学习](#)

分类号 [TP18](#)

DOI:

通讯作者:

作者个人主页: [何婷婷](#); [徐超](#); [李晶](#); [赵君喆](#)

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(92KB\)](#)

▶ [\[HTML全文\]\(OKB\)](#)

▶ [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](#)

▶ [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“命名实体”的相关文章](#)

▶ 本文作者相关文章

· [何婷婷, 徐超, 李晶, 赵君喆](#)