

[an error
occurred
while
processing
this
directive]

本期目录 | 下期目录 | 过刊浏览 | 高级检索
[关闭]

[打印本页]

论文

基于半监督学习的蛋白质关系抽取研究

崔宝今 林鸿飞 张霄

大连理工大学计算机科学与工程系, 辽宁 大连 116024

摘要:

使用半监督学习方法中的自训练、协同训练方法,利用少量已标注样本和大量未标注样本来完成蛋白质关系抽取的任务.首先使用基于词特征的SVM(support vector machine)模型进行自训练,然后使用基于词特征的SVM模型和基于依存树特征的SVM模型进行协同训练.通过对4个语料的实验,验证了自训练及协同训练方法在蛋白质关系抽取领域中的应用效果.相比于自训练,协同训练可以通过两个相对独立的视图相互补充、相互学习,进而可以有效利用未标注数据.

关键词: 关系抽取;半监督学习;词特征;自训练;协同训练

Research of protein-protein interaction extraction based on semi-supervised learning

Department of Computer Science and Engineering, Dalian University of Technology, Dalian 116024, China

Abstract:

Semi-supervised learning methods including self-training and co-training were shown in the task of PPI on how to alleviate the tag burden as much as possible. On self training a word feature based SVM model was applied; In co-training word feature and dependency tree based SVM models were used. Experiments on four copra showed that the two methods were effective in reducing the amount of labeling PPI(protein-protein interaction). Comparing self -training to cotraining can be more effective by using two separated views.

Keywords:

收稿日期 2009-05-18 修回日期 网络版发布日期

DOI:

基金项目:

国家自然科学基金资助项目(60373095,60673039);国家863高科技计划资助项目(2006AA01Z151);教育部留学回国人员科研启动基金资助项目 [教外司留(2007)1108]

通讯作者: 林鸿飞,男,内蒙古通辽人, 博士,教授,博士生导师,研究方向为搜索引擎、文本挖掘和自然语言理解. E-mail: hflin@dlut.edu.cn

作者简介:

本刊中的类似文章

扩展功能

本文信息

Supporting info

PDF(914KB)

[HTML全文]

([\\${article.html_WenJianDaXiao}](#)
KB)

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

本文关键词相关文章

关系抽取;半监督学习;词特征;自训练;协同训练

本文作者相关文章

崔宝今

林鸿飞

张霄