

计算机科学

基于隐马尔可夫模型的转录因子文本挖掘算法

吴晓洲¹, 万里明², 韩霄松¹, 梁艳春¹, 吴春国^{1,3}

1. 吉林大学 计算机科学与技术学院, 符号计算与知识工程教育部重点实验室, 长春 130012; 2. 中国人民解放军空军装备研究院 装备总体论证研究所, 北京 100076|3. 上海理工大学 管理学院, 上海 200093

摘要:

提出一种基于隐马尔可夫模型的转录因子文本挖掘算法(HMM-TFM), 该方法通过建立转录因子名称的词库, 利用谓词筛选策略判断句子是否描述转录因子, 使用隐马尔可夫模型预测单词词性, 并根据前后文单词词性识别转录因子的名称. 实验结果表明, HMM-TFM 在英文文献中抽取转录因子名称的查全率和查准率分别可达74.2%和77.9%.

关键词: 隐马尔可夫模型; 转录因子; 文本挖掘; 启动子; 生物信息

An HMM Based Transcription Factor Name Mining Algorithm

WU Xiao zhou¹, WAN Li ming², HAN Xiao song¹, LIANG Yan chun¹, WU Chun guo^{1,3}

1. College of Computer Science and Technology, Key Laboratory for Symbol Computation and Knowledge Engineering of National Education Ministry, Jilin University, Changchun 130012, China| 2. Research Institute on General Development and Evaluation of Equipment, EAAF of PLA, Beijing 100076, China; 3. School of Business, University of Shanghai for Science and Technology, Shanghai 200093, China

Abstract:

A text mining algorithm named HMM TFM (hidden Markov model based transcription factor name mining) was presented. The proposed algorithm does not need a dictionary of transcription factor names. A small verb set is defined to filter sentences. Transcription factor names are mined according to the part of speech tagged by hidden Markov model. Experimental results show that the recall rate and precision of HMM TFM come to 74.2% and 77.9%, respectively.

Keywords: hidden Markov model transcription factor text mining promoter bioinformatics

收稿日期 2011-12-29 修回日期 网络版发布日期

DOI:

基金项目:

通讯作者: 吴春国

作者简介:

作者Email: wucg@jlu.edu.cn

参考文献:

本刊中的类似文章

文章评论

扩展功能

本文信息

- ▶ Supporting info
- ▶ PDF(280KB)
- ▶ [HTML全文]
- ▶ 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

本文关键词相关文章

- 隐马尔可夫模型; 转录因子;
- ▶ 文本挖掘; 启动子; 生物信息

本文作者相关文章

- ▶ 吴晓洲
- ▶ 万里明
- ▶ 韩霄松
- ▶ 梁艳春
- ▶ 吴春国

PubMed

- ▶ Article by Tun, X. Z.
- ▶ Article by Mo, L. M.
- ▶ Article by Han, X. S.
- ▶ Article by Liang, Y. C.
- ▶ Article by Tun, C. G.

反馈人	<input type="text"/>	邮箱地址	<input type="text"/>
反馈标题	<input type="text"/>	验证码	<input type="text"/> 5977