

短文

## 一种用于基因名字规范化的多层歧义消解框架

孙承杰, 王晓龙, 林磊, 刘远超

1. 哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001

收稿日期 2007-12-12 修回日期 2008-6-13

网络版发布日期 接受日期

摘要

灵活的基因名字命名方式使基因名字具有严重的歧义, 这已成为对生物医学文献进行深层自动文本挖掘的主要障碍之一. 基因名字规范化是解决这一问题的有效途径. 本文提出了一种多层歧义消解框架来完成基因名字规范化任务. 基因名字规范化过程中不同阶段有不同的歧义情形, 在本文提出的框架中, 针对这些情形采用了有针对性的解决策略, 包括: 基于词典的基因名字检测, 基于机器学习方法的候选选择以及基于语义的歧义消解. 试验表明, 我们的方法能够在 BioCreAtIvE2006 的基因名字规范化测试集上取得 0.746 的 F 度量.

关键词

[基因名字规范化](#) [最大熵模型](#) [语义相似度](#)

分类号

## A Multi-level Disambiguation Framework for Gene Name Normalization

SUN Cheng-Jie, WANG Xiao-Long, LIN Lei, LIU Yuan-Chao

1. School of Computer Science, Harbin Institute of Technology, Harbin 150001, P.R.China

### 扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(200KB\)](#)

▶ [\[HTML全文\]\(OKB\)](#)

▶ [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](#)

▶ [复制索引](#)

▶ [Email Alert](#)

相关信息

▶ [本刊中 包含 “](#)

[基因名字规范化” 的相关文章](#)

▶ [本文作者相关文章](#)

· [孙承杰](#)

· [王晓龙](#)

· [林磊](#)

· [刘远超](#)