

人工智能及识别技术

信息检索用户查询语句的停用词过滤

熊文新<sup>1,2</sup>, 宋 柔<sup>2</sup>

(1. 北京外国语大学中国外语教育研究中心, 北京 100089; 2. 北京语言大学语言信息处理研究所, 北京 100083)

收稿日期 修回日期 网络版发布日期 2007-3-9 接受日期

**摘要** 针对以自然语言形式提出的查询请求, 区分信息需求表述和信息内容两部分。基于近20万语句的查询语料库和背景语料人民日报对照, 提出汉语通用停用词和查询专用的相对停用词, 采用左右熵和Ngram方法及KL距离脱机构造相应候选词表。根据候选词语的Bigram属性和句中不同位置的分布特点, 给出了在线动态识别停用词的方法。实验结果表明, 该方法比单纯根据静态停用词表标注效果要好。

**关键词** [用户查询](#) [停用词](#) [构造](#) [识别](#)

分类号

**DOI:**

通讯作者:

作者个人主页: [熊文新<sup>1,2</sup>; 宋 柔<sup>2</sup>](#)

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF \(121KB\)](#)
- ▶ [\[HTML全文\] \(0KB\)](#)
- ▶ [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“用户查询”的 相关文章](#)
- ▶ 本文作者相关文章
- [熊文新<sup>1,2</sup>, 宋 柔<sup>2</sup>](#)