

论文

基于互联网和self-training的中文问答模式学习

李志圣 孙越恒 何丕廉 候越先

天津大学 计算机科学与技术学院 天津大学 计算机科学与技术学院 天津大学 计算机科学与技术学院 天津大学 计算机科学与技术学院

摘要:

在已有的问答模式学习中,模式定义和候选答案评分偏于简单,而且学习过程依赖于人工标定语料。通过挖掘Web文本中动、名词序列的骨架模式,用以扩充模式定义;将self-training学习机制引入问答模式学习;用一对训练语料进行初始学习,通过互联网搜索,自动选择可靠程度较高的问答对,重新训练;扩充了启发规则,改进候选答案的评分方法。实验结果表明:所提出的问答模式学习方法能有效地提高中文问答系统的性能。

关键词: 互联网 问答模式 self-training 机器学习

Chinese question answering pattern learning based on self-training mechanism and Web

Abstract:

In the past, the learning for QA pattern relies on the labeled data, and the definition of pattern and the scoring method for the candidate answers are over simplified. The verb and noun sequence was extracted as the skeleton pattern to expand definition of QA pattern. In the learning process, a learning mechanism was established based on self-training. At first, the initial study was completed on a labeled QA pair, then the system would automatically select the reliable data for self training through searching in the Web while the system was running. The scoring method of the candidate answers was also improved by applying several heuristic rules. The experimental results show that the performance of Chinese QA system based on our method is improved significantly.

Keywords: Web QA pattern self-training machine learning

收稿日期 2007-12-14 修回日期 1900-01-01 网络版发布日期

DOI:

基金项目:

通讯作者: 李志圣

作者简介:

参考文献:

本刊中的类似文章

1. 张晗 万明杰 王寒凝 .战术互联网同质层基于信任评估的安全分簇算法[J]. 计算机应用, 2007,27(10): 2464-2469
2. 汪林林 胡德华 王佐成 宋华.基于Flex的RIA WebGIS研究与实现[J]. 计算机应用, 2008,28(12): 3257-3260

文章评论 (请注意:本站实行文责自负,请不要发表与学术无关的内容!评论内容不代表本站观点.)

扩展功能

本文信息

- ▶ Supporting info
- ▶ PDF (585KB)
- ▶ [HTML全文]
- ▶ 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

- ▶ 互联网
- ▶ 问答模式
- ▶ self-training
- ▶ 机器学习

本文作者相关文章

- ▶ 李志圣
- ▶ 孙越恒
- ▶ 何丕廉
- ▶ 候越先

PubMed

- ▶ Article by
- ▶ Article by
- ▶ Article by
- ▶ Article by

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" value="3788"/>