

论文

一种基于粗糙集增量式规则学习的问题分类方法研究

李鹏, 王晓龙, 关毅

哈尔滨工业大学计算机学院 哈尔滨 150001

收稿日期 2006-10-30 修回日期 2007-5-21 网络版发布日期 2008-7-15 接受日期

摘要

该文提出一种基于粗糙集增量式规则自动学习来实现问题分类的方法, 通过深入提取问句特征并采用决策表形式构建训练语料, 利用机器学习的方法自动获取分类规则。与其他方法相比优势在于, 用于分类的规则自动生成, 并采用粗糙集理论的简约方法获得优化的最小规则集; 首次在问题分类中引入增量式学习理念, 不但提高了分类精度, 而且避免了繁琐的重新训练过程, 大大提高了学习速度, 并且提高了分类的可扩展性和适应性。对比实验表明, 该方法分类精度高, 适应性好。在国际TREC2005Q/A实际评测中表现良好。

关键词 [粗糙集](#) [问题分类](#) [增量式学习](#) [决策表](#) [特征选择](#)

分类号 [TP391.6](#)

Question Classification with Incremental Rule Learning Algorithm Based on Rough Set

Li Peng, Wang Xiao-long, Guan Yi

School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China

Abstract

This paper presents a method on automatic question classification through incremental rule learning based on rough set theory. The core of the method is applying the machine learning approach to gain classified rules automatically through extract the features of query sentence thoroughly, and the decision table is used to construct the training collection. Comparing with the alternative means, the superiority is that it acquires the classified rule automatically and uses the rough set method to obtain the optimized smallest rule set. Especially, the incremental learning is induced to improve the precision and avoid the tedious re-training process. The performance of the approach is promising, when tested on opposite test. Meanwhile, the method obtains a very good result in the international TREC2005 Q/A track.

Key words [Rough set](#) [Question classification](#) [Incremental learning](#) [Decision table](#) [Feature selection](#)

DOI:

通讯作者

作者个人主页 李鹏; 王晓龙; 关毅

扩展功能
本文信息
▶ Supporting info
▶ PDF (259KB)
▶ [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
▶ 复制索引
▶ Email Alert
▶ 文章反馈
▶ 浏览反馈信息
相关信息
▶ 本刊中 包含“粗糙集”的 相关文章
▶ 本文作者相关文章
· 李鹏
· 王晓龙
· 关毅