论文与报告

基于多种群协同优化的文本分类规则抽取方法

刘赫, 刘大有, 裴志利, 高滢

1. 吉林大学计算机科学与技术学院 长春 130012

2. 吉林大学符号计算与知识工程教育部重点实验室 长春 130012

3. 内蒙古民族大学计算机科学与技术学院 通辽 028043

摘要

针对文本分类中的规则抽取问题, 提出一种基于多种群协同优化的文本分类规则抽取方法. 该方法利用信息熵生成初始种群, 采用多种群协同优化方法演化当前种群. 多种群协同优化方法通过种群之间的相互竞争和良种共享机制提高优化方法的效率. 实验结果表明, 本文提出的文本分类规则抽取方法所抽取规则的数量少, 准确率高, 平均长度短; 同时, 本文方法所用的计算时间少, 抽取分类规则的速度快, 适用于大规模数据集.

关键词　规则抽取　文本分类　多种群协同优化　遗传算法　蚁群算法

分类号　TP391

Rule Extraction Approach to Text Categorization Based on Multi-population Collaborative Optimization

LIU He, LIU Da-You, PEI Zhi-Li, GAO Ying

1. College of Computer Science and Technology, Jilin University, Changchun 130012

2. Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012

3. College of Computer Science and Technology, Inner Mongolia University for Nationalities, Tongliao 028043

Abstract

For the problem of rule extraction in text categorization, a novel rule extraction approach to text categorization based on multi-population collaborative optimization was proposed. Information entropy was applied to generation of initial populations and the multi-population collaborative optimization method was employed to evolve the current population in this proposed approach. The optimization efficiency of this approach was improved by the mutual competition and excellent individuals sharing mechanisms among populations. Experimental results have shown that the number of the rules extracted by this approach is small, and that the accuracy of these rules is high and the average length of them is short. Furthermore, the time of this approach is short and the speed of rule extraction through this approach is high. Therefore, this approach is suitable for large-scale data sets.

Key words　Rule extraction　text categorization　multi-population collaborative optimization　genetic algorithm　ant colony algorithm

通讯作者　刘大有 dyliu@jlu.edu.cn

作者个人主页　刘赫; 刘大有; 裴志利; 高滢

扩展功能

本文信息

▸ Supporting info
▸ PDF(1069KB)
▸ [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

相关信息

▸ 本刊中 包含"规则抽取"的 相关文章

▸ 本文作者相关文章

· 刘赫
· 刘大有
· 裴志利
· 高滢