

采用并行协同进化遗传算法的文本特征选择

邬开俊¹, 鲁怀伟²

1. 兰州交通大学 电子与信息工程学院, 兰州 730070;
2. 兰州交通大学 数理与软件工程学院, 兰州 730070

PCEGA used to solve text feature selection

WU Kai-jun¹, LU Huai-wei²

1. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;
2. School of Physics and Software Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

- 摘要
- 参考文献
- 相关文章

全文: [PDF \(755 KB\)](#) [HTML \(1 KB\)](#) 输出: [BibTeX](#) | [EndNote \(RIS\)](#) [背景资料](#)

摘要 现有的文本特征选择方法都是串行化的, 应用于海量文本数据集时时间效率较低, 因此利用并行思想来提高文本特征选择的效率, 已成为文本挖掘领域的一个研究热点. 本文将遗传算法和并行协同进化算法结合起来, 在粗糙集的基础上设计了一个并行协同进化遗传算法, 并将该算法用于文本特征选择. 该方法采用遗传算法搜索特征, 利用并行协同进化算法来提高时间效率, 从而较快地获得较具代表性的特征子集. 实验结果表明该方法是有效的.

关键词: 特征选择 文本挖掘 遗传算法 协同进化 粗糙集

Abstract: Most of existing text feature selection methods are serial and are inefficient timely to be applied to Chinese massive text data sets. So, it is a hotspot of text mining how to improve efficiency of text feature selection by means of parallel thinking. Combining genetic algorithm with parallel collaborative evolutionary, a parallel collaborative evolutionary genetic algorithm (PCEGA) based on rough sets was designed and used to select text features. The presented method took advantage of genetic algorithm to select features and employed parallel collaborative evolutionary to enhance time efficiency, so that the more representative feature subsets was acquired quickly. Experimental results show that the method is effective.

Key words: feature selection text mining genetic algorithm collaborative evolutionary rough sets




收稿日期: 2010-06-25;

基金资助: 国家自然科学基金(12CGL004); 兰州交通大学青年科学研究基金(2011005)

引用本文:

邬开俊, 鲁怀伟. 采用并行协同进化遗传算法的文本特征选择[J]. 系统工程理论实践, 2012, 32(10): 2215-2220.

WU Kai-jun, LU Huai-wei. PCEGA used to solve text feature selection[J]. Systems Engineering - Theory & Practice, 2012, 32(10): 2215-2220.

- [1] Nguyen M H, Torre F D. Optimal feature selection for support vector machines[J]. Pattern Recognition, 2010, 43(3): 584-591. 
- [2] Liu H W, Sun J G, Liu L. Feature selection with dynamic mutual information[J]. Pattern Recognition, 2009, 42(7): 1330-1339. 
- [3] Destrero A, Mosci S, Mol C D. Feature selection for high-dimensional data[J]. Computational Management Science, 2009, 6(1): 25-40. 
- [4] Xu Y. A formal study of feature selection in text categorization[J]. Journal of Communication and Computer, 2009, 6(4): 32-41.
- [5] Nandi B, Barman S, Paul S. Genetic algorithm based optimization of clustering in ad-hoc networks[J]. International Journal of Computer Science and Information Security, 2010, 7(1): 165-169.

服务

- ▶ 把本文推荐给朋友
- ▶ 加入我的书架
- ▶ 加入引用管理器
- ▶ E-mail Alert
- ▶ RSS

作者相关文章

- ▶ 邬开俊
- ▶ 鲁怀伟

- [6] Lung R I, Chira C, Dumitrescu D. An agent-based collaborative evolutionary model for multimodal optimization[C]// Proceedings of the 2008 GECCO Conference Companion on Genetic and Evolutionary Computation, USA: Atlanta, 2008: 1969-1976.
- [7] Lung R I, Dumitrescu D. A new collaborative evolutionary-swarm optimization technique[C]// Proceedings of the 2007 GECCO Conference Companion on Genetic and Evolutionary Computation, England: London, 2007: 2817-2820.
- [8] Gog A, Dumitrescu D, Hirsbrunner B. Collaborative evolutionary algorithms for combinatorial optimization[C]// Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation, England: London, 2007: 1511-1517.
- [9] 于晓义,孙树栋,褚崑.基于并行协同进化遗传算法的多协作车间计划调度[J].计算机集成制造系统, 2008, 14(5): 991-1000.Yu X Y, Sun S D, Chu W. Parallel collaborative evolutionary genetic algorithm for multi-workshop planning and scheduling problems[J]. Computer Integrated Manufacturing Systems, 2008, 14(5): 991-1000.
- [10] 谷建军. 粗糙集理论在数据约简中的应用研究[D]. 济南: 山东师范大学, 2007. Gu J J. Application of rough set theory in data reduction[D]. Jinan: Shandong Normal University, 2007. 
- [1] 李红婵, 朱颢东. 采用十进制免疫遗传算法求解高校排课问题[J]. 系统工程理论与实践, 2012, 32(9): 2031-2036.
- [2] 孙国华. 带时间窗的开放式满载车辆路径问题建模及其求解算法[J]. 系统工程理论与实践, 2012, (8): 1801-1807.
- [3] 蔡佳, 胡杰, 黄长强. 协调优势粗糙集方法及其在UCAV目标威胁估计中的应用[J]. 系统工程理论与实践, 2012, (6): 1377-1384.
- [4] 邓富民, 梁学栋, 刘爱军, 包北方. 多资源约束下改进NSGA-II算法的手术调度[J]. 系统工程理论与实践, 2012, (6): 1337-1345.
- [5] 李明伟, 康海贵, 周鹏飞. 基于NCAGA-投影寻踪混合优化城市客运量预测[J]. 系统工程理论与实践, 2012, 32(4): 903-910.
- [6] 陈敬贤, 王国华, 梁樑. 供应链系统中零售商横向转载的随机规划模型及算法[J]. 系统工程理论与实践, 2012, 32(4): 738-745.
- [7] 王新平, 王海燕. 多疫区多周期应急物资协同优化调度[J]. 系统工程理论与实践, 2012, 32(2): 283-291.
- [8] 李新然, 牟宗玉. 整数比策略下含变质产品且缺货部分需补的VMI模型[J]. 系统工程理论与实践, 2012, 32(10): 2135-2142.
- [9] 潘立军, 符卓. 求解带时间窗取送货问题的遗传算法[J]. 系统工程理论与实践, 2012, 32(1): 120-126.
- [10] 王绍仁, 马祖军. 震害紧急响应阶段应急物流系统中的LRP[J]. 系统工程理论与实践, 2011, 31(8): 1497-1507.
- [11] 黄兵, 魏大宽. 基于距离的直觉模糊粗糙模型及应用[J]. 系统工程理论与实践, 2011, 31(7): 1356-1362.
- [12] 苏国军, 汪雄海. 半导体制造系统改进Petri网模型的建立及优化调度[J]. 系统工程理论与实践, 2011, 31(7): 1372-1377.
- [13] 陈森, 姜江, 陈英武, 沈永平. 未定路网结构情况下应急物资车辆配送问题模型与应用[J]. 系统工程理论与实践, 2011, 31(5): 907-913.
- [14] 孙杨, 宋瑞, 何世伟. 随机需求下公交时刻表设计的鲁棒性优化[J]. 系统工程理论与实践, 2011, 31(5): 986-992.
- [15] 杨丰梅, 任姝仪, 周荣喜. 带有惩罚项的多项式样条函数利率期限结构模型实证比较[J]. 系统工程理论与实践, 2011, 31(4): 735-739.

版权所有 © 2011 《系统工程理论与实践》编辑部

地址: 北京中关村东路55号 100190 电话: 010-62541828 Email: xtl@chinajournal.net.cn

本系统由北京玛格泰克科技发展有限公司设计开发 技术支持: support@magtech.com.cn