

数据库、信息处理

使用PGA的特征选择方法

马春华¹, 朱颢东²

1.绥化学院 计算机科学与技术系, 黑龙江 绥化 152061

2.中国科学院 成都计算机应用研究所, 成都 610041

收稿日期 2009-4-8 修回日期 2009-6-5 网络版发布日期 接受日期

摘要 特征选择是文本分类系统的核心步骤之一。然而现有的特征选择方法都是串行化的, 应用于中文海量文本数据时时间效率较低, 因此利用并行策略来提高特征选择的效率, 已经成为研究的热点。详细设计了一个用于特征选择的并行遗传算法, 该算法采用遗传算法搜索特征, 利用并行策略评价特征子集, 即将种群中个体的适应度计算并行在多个计算节点上同时进行, 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用的。

关键词 [文本分类](#) [特征选择](#) [遗传算法](#) [并行策略](#)

分类号

Feature selection method applied PGA

MA Chun-hua¹, ZHU Hao-dong²

1.Department of Computer Science and Technology, Suihua College, Suihua, Heilongjiang 152061, China

2.Chengdu Institute of Computer Application, Chinese Academy of Sciences, Chengdu 610041, China

Abstract

Feature selection is one of the key steps in text classification system. However, most of existing feature selection methods are serial and are inefficient timely to be applied to Chinese massive text data sets, so it is a hotspot how to improve efficiency of feature selection by means of parallel strategy. It detailedly designs a Parallel Genetic Algorithm (PGA) which is used to select features. The algorithm uses genetic algorithm to search features and calculates fitness of feature subsets in multiple computing nodes at the same time, so can acquire quickly feature subsets which are more representative. Experimental results show that the method is effective.

Key words [text categorization](#) [feature selection](#) [Genetic Algorithm \(GA\)](#) [parallel strategy](#)

DOI: 10.3778/j.issn.1002-8331.2009.22.035

通讯作者 马春华 shzlm009@sina.com

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(535KB\)](#)

▶ [HTML全文\(0KB\)](#)

▶ [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“文本分类”的相关文章](#)

▶ 本文作者相关文章

· [马春华](#)

· [朱颢东](#)