山东大学学报(工学版) 2012, 42(4) 19-23 DOI: ISSN: 0412-1961 CN: 21-1139

机器学习与数据挖掘

基于云计算平台的代价敏感集成学习算法研究

张伶卫,万文强

南京邮电大学计算机学院, 江苏 南京 210003

摘要:

针对现实生活中大规模不平衡数据的分类问题,设计了一种基于云计算平台的代价敏感集成学习分类算法。Hadoop云计算平台对海量数据进行划分用于并行学习,同时结合代价敏感的思想对学习得到的基分类器进行加权集成,实现了云计算平台上的代价敏感集成学习分类模型。仿真实验表明该模型能够明显提高少数类的查全率,同时Hadoop的并行机制使得云平台环境下的集成学习时间较集中式环境有大幅度的缩减,进一步提高了大规模不平衡数据分类问题的学习效率。

关键词: 代价敏感 集成学习 云计算平台 不平衡分类; 分布式

Study on the cost-sensitive ensemble learning algorithm based on the cloud computing platform

ZHANG Ling-wei, WAN Wen-qiang

College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Abstract:

With respect to the classification of large scale imbalanced data, a distributed cost-sensitive ensemble learning algorithm based on cloud computing platform was proposed. The large scale data was divided on Hadoop cloud computing platform and was used in parallel learning. Based on the idea of cost-sensitive, a weighted ensemble classifier was achieved, and a distributed cost-sensitive ensemble learning model based on cloud computing platform was developed. Experiment results showed that the recall rate of the minority class was improved significantly and the computational time was shortened by the ensemble learning on cloud computing platform due to the Hadoop parallel mechanism. In additron, the classification efficiency of the large-scale imbalanced problem was largely improved.

Keywords: cost sensitive learning ensemble learning cloud computing platform imbalanced pattern classification distribution

收稿日期 2012-05-15 修回日期 网络版发布日期

DOI:

基金项目:

国家重点基础研究发展计划(973计划)资助项目(2011CB302903);国家自然科学基金资助项目(61073114);南京邮电大学攀登计划资助项目(NY210010)

通讯作者:

作者简介: 张伶卫(1989-),男,江苏南京人,硕士研究生,主要研究方向为数据挖掘与机器学习. E-mail:1010041215@njupt.edu.cn

作者Email:

PDF Preview

参考文献:

本刊中的类似文章

1. 李霞1,王连喜2,蒋盛益1.面向不平衡问题的集成特征选择[J]. 山东大学学报(工学版), 2011,41(3): 7-11

2. 李小斌1, 李世银2.时间序列早期分类的多分类器集成方法[J]. 山东大学学报(工学版), 2011,41(4): 73-78

3．谢伙生,刘敏.一种基于主动学习的集成协同训练算法[J]. 山东大学学报(工学版), 2012,42(3): 1-5
4．房晓南1,2,张化祥1,2*,高爽1,2.基于SMOTE和随机森林的Web spam检测[J]. 山东大学学报(工学版), 2013,43(1): 22-27
