

数据库、信号与信息处理

数据集动态重构的集成迁移学习

刘伟, 张化祥

山东师范大学 信息科学与工程学院, 济南 250014

收稿日期 2008-12-23 修回日期 2009-2-27 网络版发布日期 2010-4-21 接受日期

摘要 目前很多数据挖掘和机器学习方法都有一个基本假设: 训练数据和测试数据必须服从相同的分布。但是在很多情况下这种假设不成立, 没有考虑分布差异的传统机器学习方法就不能正确分类了。提出了一种新的迁移学习方法DRTAT, 对原训练数据进行动态分割重组, 适时地淘汰冗余数据, 并进行分类器的集成。通过在多个文本数据集和UCI数据集上进行测试, 并与TrAdaboost算法进行比较, 表明了算法的先进性。

关键词 [分布差异](#) [知识迁移](#) [动态数据集重组](#) [冗余数据淘汰](#) [分类器集成](#)

分类号 [TP391.4](#)

Ensemble transfer learning algorithm based on dynamic dataset regroup

LIU Wei, ZHANG Hua-xiang

College of Information Science and Engineering, Shandong Normal University, Jinan 250014, China

Abstract

There is a basic assumption in many existing data mining and machine learning techniques, that training and test data must be governed by the same distribution. However, this assumption does not hold in many cases, then traditional machine learning methods not aware of the difference of distribution may fail. This paper proposes a novel transfer-learning algorithm called DRTAT, which dynamically regroups the primary training data sets and eliminates the redundancy data timely, then makes classifiers ensemble. The experiments are performed on many text data sets and the UCI benchmark data sets, and DRTAT is compared with TrAdaboost algorithm, the results show the superiority of DRTAT.

Key words [distribution difference](#) [knowledge transfer](#) [dynamic dataset regroup](#) [eliminating the redundancy data](#) [classifier ensemble](#)

DOI: 10.3778/j.issn.1002-8331.2010.12.037

通讯作者 刘伟 wei4480@163.com

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(536KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [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“分布差异”的相关文章](#)
- ▶ [本文作者相关文章](#)

- [刘伟](#)
- [张化祥](#)