

数据库、信号与信息处理

SVM中不平衡数据的分离超平面的校正方法

刘万里, 刘三阳

洛阳师范学院 数学科学学院, 河南 洛阳 471022

收稿日期 2007-9-28 修回日期 2008-1-21 网络版发布日期 2008-6-26 接受日期

摘要 针对两类不平衡数据的分离超平面的偏移问题提出一种平衡方法。首先, 对两类样本数据在核空间中进行核主成分分析, 分别求出两类样本数据在特征空间中的主要特征值; 然后, 根据两样本容量以及各自的特征值所提供的信息, 对两类数据给出惩罚因子比例; 最后, 通过优化训练, 产生一个新的分离超平面。该分类面校正了标准的支持向量机的分类误差。实验显示了所提出方法的有效性, 即与标准的支持向量机相比, 不仅平衡了错分率而且还能减少错分率。

关键词 [不平衡数据](#) [核主成分分析](#) [支持向量机](#) [偏移](#)

分类号

Revising method for separation hyperplane of imbalanced data in SVM

LIU Wan-li, LIU San-yang

Department of Mathematics, Luoyang Normal College, Luoyang, Henan 471022, China

Abstract

A balance method for the offset of separation hyperplane of biclassification imbalanced data is proposed. Firstly, the principal eigenvalues are found respectively of the two classes of samples in feature space by using Kernel Principal Component Analysis (KPCA). Secondly, one penalty proportion is given based on the information provided by the sizes of the two sample data and their eigenvalues. Finally, a new separation hyperplane is generated by the optimization training. The hyperplane revises the error of the standard Support Vector Machines. Experiments show the efficiency of proposed method, i.e. comparing with standard Support Vector Machines the classification error can be balanced and be also decreased.

Key words [imbalanced data](#) [Kernel Principal Component Analysis \(KPCA\)](#) [Support Vector Machines \(SVM\)](#) [offset](#)

DOI:

通讯作者 刘万里 lwali@lynu.edu.cn

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF\(515KB\)](#)

▶ [\[HTML全文\]\(0KB\)](#)

▶ [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“不平衡数据”的相关文章](#)

▶ [本文作者相关文章](#)

· [刘万里](#)

· [刘三阳](#)