

# 基于噪声概率的多核集成算法

冯 杰, 武德安, 吴 磊

(电子科技大学 数学科学学院, 四川 成都 611731)

**摘 要:** 为解决多核集成 (Multiple Kernel Boosting, MKBoost) 算法对噪声敏感的问题, 考虑理想分类器区别对待正常样本与噪声样本的特点, 提出了一种适用于被噪声污染数据集的多核集成算法. 采用 KNN(K 最近邻)方法与 logistic 回归的融合, 构造了样本噪声概率函数, 计算出每个样本是噪声的概率, 根据噪声概率构造了新的损失函数, 利用加法模型得到每轮迭代的基分类器系数. UCI 数据集上的实验结果表明, 该算法可以有效降低多核集成算法对噪声的敏感程度, 提高了鲁棒性.

**关键词:** 多核学习; 提升方法; K 最近邻法; logistic 回归; 支持向量机; 鲁棒性

## A Noise-Probability Based Multiple Kernel Boosting Algorithm

FENG Jie, WU De-an, WU Lei

(School of Mathematics, University of Electronic Science and Technology of China, Chengdu 611731, China)

**Abstract:** To solve the problem that the Multiple Kernel Boosting (MKBoost) algorithm was sensitive to noise, the characteristic that noise data and normal data were treated differently by the optimal classifier was considered, and a MKBoost algorithm applied to datasets polluted by noise data was proposed. The fuse of KNN (K-nearest neighbor) algorithm and logistic regression was used for the construction of noise probability function, and calculating the probability of each instance being noise. A new loss function was constructed via the noise probability function, and additive model was utilized for computing the coefficients corresponded to base classifier in each iteration. Experiments on the datasets from UCI show that the proposed method could efficiently reduce the sensitivity of MKBoost algorithm to noise, and improved the robustness

**Key words:** multiple kernel learning; boosting method; KNN; logistic regression; SVM; robustness

**作者简介:**

冯 杰 男, (1989-), 硕士研究生. 研究方向为数据挖掘与计算统计.

E-mail: 1661798458@qq.com.

武德安 男, (1971-), 博士, 副教授. 研究方向为模式识别、数据挖掘.

吴 磊 男, (1978-), 博士, 副教授. 研究方向为模式识别、数据挖掘.