论文与报告

Kernel-*k*NN: 基于信息能度量的核*k*-最近邻算法

刘松华, 张军英, 许进, 贾宏恩

1. 西安电子科技大学计算机学院 西安 710071

2. 北京大学信息科学技术学院 北京 100871

3. 西安交通大学理学院 西安 710049

摘要

提出一种核*k*最近邻算法. 首先给出用于最近邻学习的信息能度量方法, 该方法克服了高维数据不便于用传统距离度量表示的困难, 提高了数据间类别相似性和距离的一致性. 在此基础上, 将传统的*k*NN扩展为非线性形式, 并采用半正定规划学习全局最优的度量矩阵. 算法主要特点是: 能较好地适用于高维数据, 并有效提升*k*NN 的分类性能. 多个数据集的实验和分析表明, 本文的Kernel-*k*NN算法与传统的*k*NN算法比较, 在低维数据上, 分类准确率相当; 在高维数据上, 分类性能有明显提高.

关键词　距离度量　非线性变换　*k*-最近邻(*k*-NN)　核方法

分类号

Kernel-*k*NN: A New *k*NN Algorithm Based on Informational Energy Metric

LIU Song-Hua, ZHANG Jun-Ying, XU Jin, JIA Hong-En

1. School of Computer Science and Technology, Xidian University, Xi'an 710071

2. School of Electronics Engineering and Computer Science, Peking University, Beijing 100871

3. College of Science, Xi'an Jiaotong University, Xi'an 710049

Abstract

This paper proposes a new algorithm named Kernel-*k*NN. To begin with, an approach for information energy metric is proposed, which is used to learn the nearest neighbor. This method overcomes the inconvenience for distance metric expression with high dimensional data set, and improves the consistency between the class similarity and the distance. Meanwhile, the traditional *k*NN is extended to an nonlinear form, and semidefinite programming is usd to learn the globally optimal metric matrix. The main characteristic of the proposed algorithm is that it is suitable for high dimensional data set, and can improve the classification performance efficiently. Experiments and analysis on many data sets have shown that Kernel-*k*NN can get the common performance in low dimensional data, and have a significant improvement on large scale data in high dimensions.

Key words　Distance metric　nonlinear transformation　*k*-nearest neighbor (*k*NN)　kernel method

通讯作者　刘松华 sooh.liu@gmail.com

作者个人主页　刘松华; 张军英; 许进; 贾宏恩

扩展功能

本文信息

▶ Supporting info
▶ PDF(2091KB)
▶ [HTML全文](0KB)
▶ 参考文献[PDF]
▶ 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

相关信息

▶ 本刊中 包含"距离度量"的 相关文章

▶本文作者相关文章

· 刘松华
· 张军英
· 许进
· 贾宏恩