

一种基于局部结构保持的数据降维方法

张琳

(遵义师范学院 计算机与信息科学学院, 贵州 遵义 563002)

摘要: 针对传统线性回归方法进行数据降维, 不能很好地利用数据间几何结构的问题, 提出了一种基于局部结构保持的数据降维方法. 利用非负约束下的最小二乘构造关系图, 从而描述局部近邻几何信息. 在将全局结构信息与局部结构信息整合时, 设计了一种新的模型选择方法用于模型参数估计, 极大地减少了计算量. 在公测库 ORL、YaleB、USPS、20Newgroup、Isolet 上的实验结果表明, 该方法用于数据分类时优于 ULDA、OLDA 和 NPE 等方法.

关键词: 数据降维; 几何结构; 回归模型

A Local Structure Preserving Based Data Dimensionality

Reduction Approach

ZHANG Lin

(Faculty of Computer and Information Science, Zunyi Normal College, Guizhou 563002, China)

Abstract: The traditional linear regression method used for data dimensionality reduction failed to capture the geometry structure of data. In order to solve this problem, we present a local structure preserving based data dimensionality reduction. A nonnegative least squares method is introduced to construct the graph, which describe the local neighborhood geometry structure information. For the integrating of global structural information and local structure information integration, we also introduce a new model selection method for model parameter estimation, greatly reducing the cost of computation. Experiments on five benchmark datasets (ORL, YaleB, USPS, 20Newgroup and Isolet) show that the proposed approach can achieve better results than ULDA, OLDA and NPE.

Key words: data dimensionality reduction; geometric structure; regression model

作者简介:

张琳女, (1985-), 硕士, 讲师. 研究方向为数据挖掘技术、计算机教育. E-mail: zyzhanglin@163.com