论文

一种优化的基于LPI的文档索引算法FLPI

陶剑文 赵杰煜

浙江工商职业技术学院 信息工程学院;宁波大学 信息与工程学院 宁波大学 信息与工程学院

摘要：

LPI对于局部流形结构是优化的，但在时空上运行效率较低，使其很难应用于大型数据集。基于LPI算法，提出了一种优化的LPI算法FLPI，它将LPI问题分解为一个图嵌入问题和一个正则最小二乘问题，避免了稠密矩阵的特征值分解，显著减少了计算复杂度。此外，在监督环境下，利用一个特别设计的图，使FLPI只需要解决正则最小二乘问题，进一步减少了时空开销。实时数据集实验结果显示，FLPI获得了相似或优于LPI的结果，且运行速度明显提升。

关键词： 局部保留索引　潜在语意索引　文档索引　维度归约

FLPI：An optimal algorithm for document indexing based on LPI

Abstract:

LPI is optimal in the sense of local manifold structure. However, LPI is not efficient in time and memory, which makes it difficult to be applied to very large data set. Therefore, an optimal algorithm called FLPI was proposed. FLPI decomposed the LPI problem into a graph embedding problem plus a regularized least squares problem. Such modification avoids eigen decomposition of dense matrices and can significantly reduce both time and memory cost in computation. Moreover, with a specifically designed graph in supervised situation, LPI only needs to solve the regularized least squares problem which is a further saving of time and memory. Experimental results on real data show that FLPI obtains similar or better results compared to LPI and it is significantly faster.

Keywords: Locality Preserving Indexing (LPI)　Latent Semantic Indexing(LSI)　document indexing dimensionality reduction

通讯作者：陶剑文

作者简介：

参考文献：

本刊中的类似文章

文章评论 (请注意:本站实行文责自负, 请不要发表与学术无关的内容!评论内容不代表本站观点.)

反馈人		邮箱地址	
反馈标		验证码	6148

扩展功能

本文信息

▶ Supporting info
▶ PDF(682KB)
▶ [HTML全文]
▶ 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
▶ 引用本文
▶ Email Alert
▶ 文章反馈
▶ 浏览反馈信息

本文关键词相关文章

▶ 局部保留索引
▶ 潜在语意索引
▶ 文档索引
▶ 维度归约

本文作者相关文章

▶ 陶剑文
▶ 赵杰煜

PubMed

▶ Article by
▶ Article by