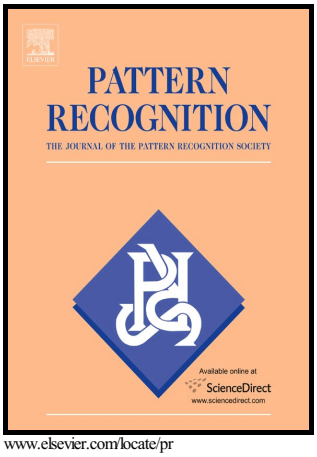

www.elsevier.com/locate/pr

Global and local structure preserving sparse subspace learning: an iterative approach to unsupervised feature selection

Nan Zhou, Yangyang Xu, Hong Cheng, Jun Fang, Witold Pedrycz

PII: S0031-3203(15)00467-7
DOI: http://dx.doi.org/10.1016/j.patcog.2015.12.008
Reference: PR5593

To appear in: *Pattern Recognition*

Received date: 15 June 2015
Revised date: 20 October 2015
Accepted date: 16 December 2015

Cite this article as: Nan Zhou, Yangyang Xu, Hong Cheng, Jun Fang and Witolc Pedrycz, Global and local structure preserving sparse subspace learning: a iterative approach to unsupervised feature selection, *Pattern Recognitior* http://dx.doi.org/10.1016/j.patcog.2015.12.008

Global and Local Structure Preserving Sparse Subspace Learning: An Iterative Approach to Unsupervised Feature Selection

Nan Zhou[a,b], Yangyang Xu[c], Hong Cheng[a,], Jun Fang[a], Witold Pedrycz[b]

[a]*Center for Robotics, School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 611731, China*
[b]*Department of Electrical and Computer Engineering, University of Alberta, AB, T6G2R3, Canada*
[c]*Department of Combinatorics and Optimization, University of Waterloo, ON, N2L3G1, Canada*

Abstract

As we aim at alleviating the curse of high-dimensionality, subspace learning is becoming more popular. Existing approaches use either information about global or local structure of the data, and few studies simultaneously focus on global and local structures as the both of them contain important information. In this paper, we propose a global and local structure preserving sparse subspace learning (GLoSS) model for unsupervised feature selection. The model can simultaneously realize feature selection and subspace learning. In addition, we develop a greedy algorithm to establish a generic combinatorial model, and an iterative strategy based on an accelerated block coordinate descent is used to solve the GLoSS problem. We also provide whole iterate sequence convergence analysis of the proposed iterative algorithm. Extensive experiments are conducted on real-world datasets to show the superiority of the proposed approach over several state-of-the-art unsupervised feature selection approaches.

Keywords: Machine learning, Feature selection, Subspace learning, Unsupervised learning

1. Introduction

With the advances in data processing, the dimensionality of the data increases and can be extremely high in many fields such as computer vision, machine learning and image processing. The high dimensionality of the data not only greatly increases the time and storage space required to realize data analysis but also introduces much redundancy and noise which can decrease the accuracy of ensuing methods. Hence, dimensionality reduction becomes an important and often necessary preprocessing step to accomplish certain machine learning tasks such as clustering and classification.

Email addresses: nzhouuestc@gmail.com (Nan Zhou), yangyang.xu@uwaterloo.ca (Yangyang Xu), hcheng@uestc.edu.cn (Hong Cheng), JunFang@uestc.edu.cn (Jun Fang), wpedrycz@ualberta.ca (Witold Pedrycz)