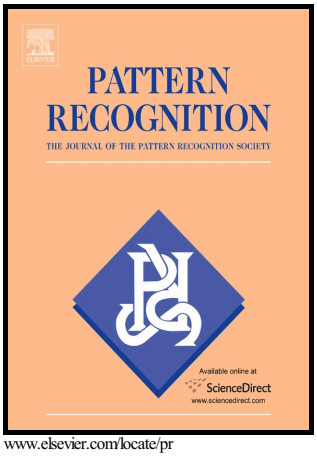
Subspace Clustering guided Unsupervised Feature Selection

Pengfei Zhu, Wencheng Zhu, Qinghua Hu, Changqing Zhang, Wangmeng Zuo



www.elsevier.com/locate/pr

PII: S0031-3203(17)30015-8
DOI: http://dx.doi.org/10.1016/j.patcog.2017.01.016
Reference: PR6018

To appear in: *Pattern Recognition*

Received date: 22 June 2016
Revised date: 4 January 2017
Accepted date: 9 January 2017

Cite this article as: Pengfei Zhu, Wencheng Zhu, Qinghua Hu, Changqing Zhang and Wangmeng Zuo, Subspace Clustering guided Unsupervised Feature Selection, *Pattern Recognition,* http://dx.doi.org/10.1016/j.patcog.2017.01.016

Subspace Clustering guided Unsupervised Feature Selection

Pengfei Zhu[a], Wencheng Zhu[a], Qinghua Hu[a], Changqing Zhang[a], Wangmeng Zuo[b]

[a]*School of Computer Science and Technology, Tianjin Univerity, Tianjin 300350, China*
[b]*School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China*

Abstract

Unsupervised feature selection (UFS) aims to reduce the time complexity and storage burden, improve the generalization ability of learning machines by removing the redundant, irrelevant and noisy features. Due to the lack of training labels, most existing UFS methods generate the pseudo labels by spectral clustering, matrix factorization or dictionary learning, and convert UFS to a supervised problem. The learned clustering labels reflect the data distribution with respect to classes and therefore are vital to the UFS performance. In this paper, we proposed a novel subspace clustering guided unsupervised feature selection (SCUFS) method. The clustering labels of the training samples are learned by representation based subspace clustering, and features that can well preserve the cluster labels are selected. SCUFS can well learn the data distribution in that it uncovers the underlying multi-subspace structure of the data and iteratively learns the similarity matrix and clustering labels. Experimental results on benchmark datasets for unsupervised feature selection show that SCUFS outperforms the state-of-the-art UFS methods.