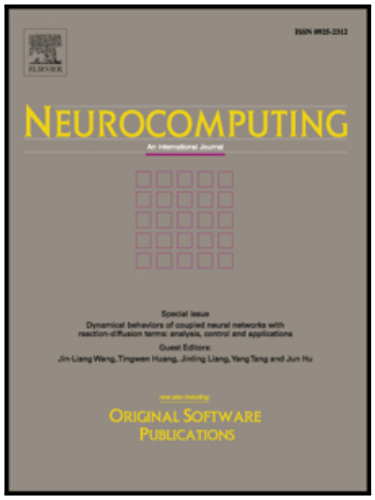
Communicated by Feiping Nie

Accepted Manuscript



Low-Rank Structure Preserving for Unsupervised Feature Selection

Wei Zheng, Chunyan Xu, Jian Yang, Junbin Gao, Fa Zhu

PII: S0925-2312(18)30746-X
DOI: 10.1016/j.neucom.2018.06.010
Reference: NEUCOM 19688

To appear in: *Neurocomputing*

Received date: 5 September 2017
Revised date: 7 April 2018
Accepted date: 19 June 2018

Please cite this article as: Wei Zheng, Chunyan Xu, Jian Yang, Junbin Gao, Fa Zhu, Low-Rank Structure Preserving for Unsupervised Feature Selection, *Neurocomputing* (2018), doi: 10.1016/j.neucom.2018.06.010

Low-Rank Structure Preserving for Unsupervised Feature Selection

Wei Zheng[a,b], Chunyan Xu[a,*], Jian Yang[a], Junbin Gao[c], Fa Zhu[a]

[a] *School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing,210094, P.R. China*
[b] *School of Computer Engineering, Jinling Institute of Technology, Nanjing,211169, P.R. China*
[c] *Discipline of Business Analytics, University of Sydney Business School, Universtiy of Sydney, NSW 2006, Australia*

Abstract

Unsupervised feature selection has been widely applied to machine learning and pattern recognition, as it does not require class labels. The majority of the popular unsupervised feature selection methods focus on various forms of reconstruction, and minimize the reconstruction residual by discarding features with low contributions. However, they cannot effectively preserve the data distribution in multiple subspaces, because the sample structure information is not substantially utilized to constrain the selected features. In this paper, we propose a low-rank structure preserving method for unsupervised feature selection (LRPFS) to address this shortcoming. The data matrix consisting selected features is assumed as a dictionary, which is learned by a low-rank constraint to preserve the subspace structure. Meanwhile, we further leverage the sparse penalty to remove the redundancy features, and thus obtain the discriminative features with intrinsic structures. In this way, the sample distribution can be preserved by low-rank constraint more precisely via using discriminative features. In turn, the refined sample structure boosts the selection of more representative features. The effectiveness of our method is supported by both theoretical and experimental results.

Keywords: Low-rank, Unsupervised learning, Feature selection, Dictionary learning

1. Introduction

Data acquisition becomes progressively convenient with the rapid development of computer hardware and Internet. In many practical areas, such as face

*Corresponding author
Email addresses: vividzheng@163.com (Wei Zheng), cyx@njust.edu.cn (Chunyan Xu), csjyang@njust.edu.cn (Jian Yang), junbin.gao@sydney.edu.au (Junbin Gao), zhufag@gmail.com (Fa Zhu)