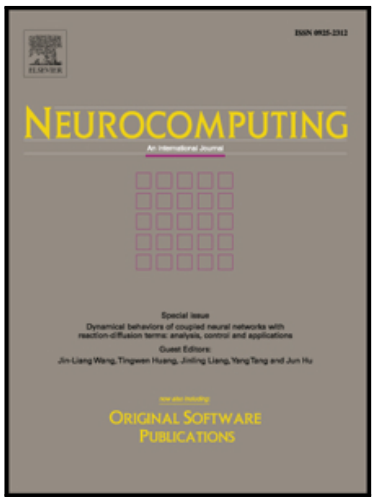

Unsupervised feature selection by regularized matrix factorization

Miao Qi , Ting Wang , Fucong Liu , Baoxue Zhang , Jianzhong Wang , Yugen Yi

PII: S0925-2312(17)31468-6
DOI: 10.1016/j.neucom.2017.08.047
Reference: NEUCOM 18839

To appear in: *Neurocomputing*

Received date: 21 May 2016
Revised date: 28 April 2017
Accepted date: 18 August 2017

Please cite this article as: Miao Qi , Ting Wang , Fucong Liu , Baoxue Zhang , Jianzhong Wang , Yugen Yi , Unsupervised feature selection by regularized matrix factorization, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.08.047

Unsupervised feature selection by regularized matrix factorization

Miao Qi[1], Ting Wang[1, 5], Fucong Liu[1], Baoxue Zhang[3], Jianzhong Wang[1, 4]* and Yugen Yi[2,]*

[1]School of Computer Science and Information Technology, Northeast Normal University, Changchun, China

[2]School of Software, Jiangxi Normal University, Nanchang, China

[3]School of Statistics, Capital University of Economics and Business, Beijing, China

[4]Key Laboratory of Applied Statistics of MOE, China

[5]School of Computer Science and Engineering, South China University of Technology, Guangzhou, China

*Corresponding authors: Jianzhong Wang and YugenYi{wangjz019, yiyg510@nenu.edu.cn}

Abstract Feature selection is an interesting and challenging task in data analysis process. In this paper, a novel algorithm named Regularized Matrix Factorization Feature Selection (RMFFS) is proposed for unsupervised feature selection. Compared with other matrix factorization based feature selection methods, a main advantage of our algorithm is that it takes the correlation among features into consideration. Through introducing an inner product regularization into our algorithm, the features selected by RMFFS would not only well represent the original high-dimensional data, but also contain low redundancy. Moreover, a simple yet efficient iteratively updating algorithm is also developed to solve the proposed RMFFS. Extensive experimental results on nine real world databases demonstrate that our proposed method can achieve better performance than some state-of-the-art unsupervised feature selection methods.

Keywords: Dimensionality reduction; Feature selection; Matrix factorization; Sparsity and redundancy

1. Introduction

Nowadays, the data obtained in many real-world applications such as pattern recognition, computer vision and image processing is often high-dimensional. High-dimensional data not only makes the model learning process to be