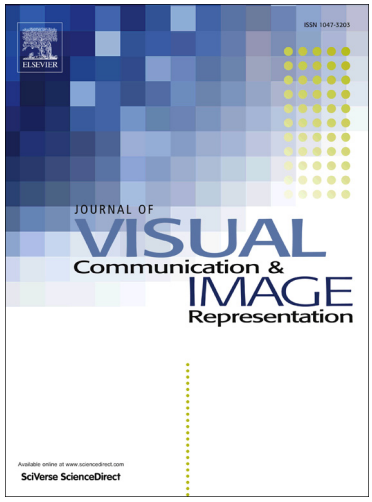

Semi-supervised Feature Selection with Exploiting Shared Information Among Multiple Tasks

Xiao-dong Wang, Rung-Ching Chen, Fei Yan, Zhi-qiang Zeng

PII: S1047-3203(16)30212-7
DOI: http://dx.doi.org/10.1016/j.jvcir.2016.10.007
Reference: YJVCI 1875

To appear in: *J. Vis. Commun. Image R.*

Received Date: 1 April 2016
Revised Date: 22 September 2016
Accepted Date: 12 October 2016

Please cite this article as: X-d. Wang, R-C. Chen, F. Yan, Z-q. Zeng, Semi-supervised Feature Selection with Exploiting Shared Information Among Multiple Tasks, *J. Vis. Commun. Image R.* (2016), doi: http://dx.doi.org/10.1016/j.jvcir.2016.10.007

Semi-supervised Feature Selection with Exploiting Shared Information Among Multiple Tasks

Xiao-dong Wang[a,b,*], Rung-Ching Chen[b,*], Fei Yan[a], Zhi-qiang Zeng[a]

[a] *College of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China*
[b] *Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan*

Abstract

Given several related tasks, multi-task feature selection determines the importance of features by mining the correlations between them. There have already many efforts been made on the supervised multi-task feature selection. However, in real-world applications, it's noticeably time-consuming and unpractical to collect sufficient labeled training data for each task. In this paper, we propose a novel feature selection algorithm, which integrates the semi-supervised learning and multi-task learning into a joint framework. Both the labeled and unlabeled samples are sufficiently utilized for each task, and the shared information between different tasks is simultaneously explored to facilitate decision making. Since the proposed objective function is non-smooth and difficult to be solved, we also design an efficient iterative algorithm to optimize it. Experimental results on different applications demonstrate the effectiveness of our algorithm.

Keywords: semi-supervised learning, feature selection, multi-task learning, Face recognition, 3D motion data analysis, Spoken Letter Recognition, Handwritten Digits Recognition

1. Introduction

High dimensional data is constantly confronted in many areas, such as pattern recognition, computer vision and multimedia analysis[1]. It is significantly time consuming to directly deal with these high dimensional data[2]. Moreover, various leaning tasks, such as classification and clustering, are prone to lose their efficiency in processing huge number of features. Recent works have shown that the most important information is carried in a subset of features, and if such feature subset is determined, higher accuracy will be obtained[3, 4].

Feature selection, which is one of the most effective methods to reduce the dimension of high dimensional data, currently renews the researcher's interest. In the past decades, lots of feature selection algorithms have been proposed and demonstrated their performance in many applications. Existing feature selection algorithms can be roughly divided into two classes, i.e., supervised feature selection and unsupervised feature selection. Supervised feature selection algorithms usually adopt the labeled training data to evaluate the feature relevance. Given sufficient label information, supervised algorithms are usually able to select the discriminative features. However, in real-world applications, collecting large number of labeled data is not practical and requires especially expensive human labor. Meanwhile, it is much easier to obtain unlabeled data, which could be effectively utilized by unsupervised feature selection[5, 6]. The most obvious example of unsupervised feature selection algorithms, e.g., Data variance[5], evaluates the features by the variance along a dimension, and the features with maximum variance will be selected. Laplacian score[6], which can be regarded as the extension of Data Variance, not only chooses features by the largest variance, but also considers the fact that the local structure of the data space is more important than the global one and selects features best preserving locality power. However, these methods neglect the inter-dependency between features when multiple features need to be selected. Recent years, a popular used criterion is to select the representative features by utilizing the manifold structure on the whole feature set[7, 8, 9]. Nevertheless, in unsupervised scenarios, there is no label information available, making it difficult to select discriminative features.

*Corresponding author.
Email addresses: xdwangjsj@xmut.edu.cn (Xiao-dong Wang), crching@cyut.edu.tw (Rung-Ching Chen)