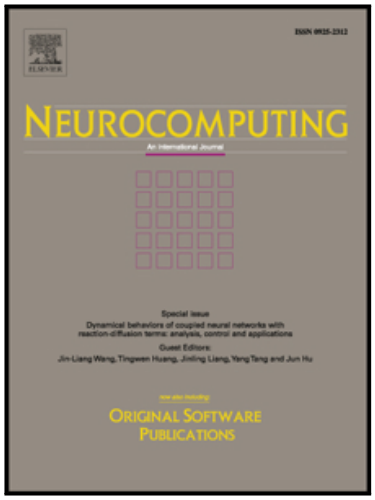
Communicated by Feiping Nie

Accepted Manuscript



Distribution Preserving Learning for Unsupervised Feature Selection

Ting Xie, Pengfei Ren, Taiping Zhang, Yuan Yan Tang

PII: S0925-2312(18)30167-X
DOI: 10.1016/j.neucom.2018.02.032
Reference: NEUCOM 19322

To appear in: *Neurocomputing*

Received date: 26 August 2016
Revised date: 19 June 2017
Accepted date: 6 February 2018

Please cite this article as: Ting Xie, Pengfei Ren, Taiping Zhang, Yuan Yan Tang, Distribution Preserving Learning for Unsupervised Feature Selection, *Neurocomputing* (2018), doi: 10.1016/j.neucom.2018.02.032

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Distribution Preserving Learning for Unsupervised Feature Selection

Ting Xie[a], Pengfei Ren[b], Taiping Zhang[b,*], Yuan Yan Tang[c]

[a] *College of of Mathematics and Statistics, Chongqing University, Chongqing, China, 400030*
[b] *College of Computer Science, Chongqing University, Chongqing, China, 400030*
[c] *Faculty of Science and Technology, University of Macau, Macau, China*

Abstract

Selection of most relevant features from high-dimensional data is difficult especially in unsupervised learning scenario, this is because there is an absence of class labels that would guide the search for relevant features. In this work, we propose a distribution preserving feature selection (DPFS) method for unsupervised feature selection. Specifically, we select those features such that the distribution of the data can be preserved. Theoretical analysis show that our proposed DPFS method share some excellent properties of kernel method. Moreover, traditional "wrapper" and "filter" feature selection methods often involve an exhaustive search optimization, feature selection problem is treated as variable of optimization problem in our proposed method, the optimization is tractable. Extensive experimental results over various real-life data sets have demonstrated the effectiveness of the proposed algorithm.

Keywords: Feature selection, Density preserving, Kernel density estimation, Dimensionality reduction, Data mining.

1. Introduction

many pattern recognition and machine learning applications, high-dimensional data becomes very common. High dimensionality not only significantly increases

*Corresponding author, Tel: 86-23-65112784, Fax: 86-23-65102502
Email address: tpzhang@cqu.edu.cn (Taiping Zhang)

Download English Version:

https://daneshyari.com/en/article/6864337

Download Persian Version:

https://daneshyari.com/article/6864337

Daneshyari.com