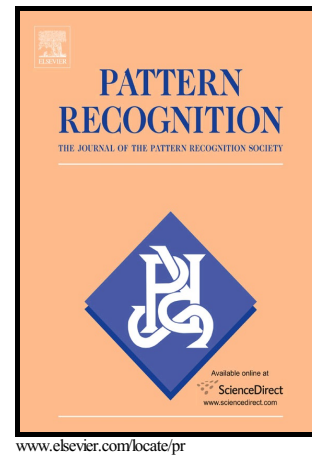


# Author's Accepted Manuscript

An efficient semi-supervised representatives feature selection algorithm based on information theory

Yintong Wang, Jiandong Wang, Hao Liao, Haiyan Chen



PII: S0031-3203(16)30224-2  
DOI: <http://dx.doi.org/10.1016/j.patcog.2016.08.011>  
Reference: PR5840

To appear in: *Pattern Recognition*

Received date: 18 January 2015  
Revised date: 24 May 2016  
Accepted date: 12 August 2016

Cite this article as: Yintong Wang, Jiandong Wang, Hao Liao and Haiyan Chen, An efficient semi-supervised representatives feature selection algorithm based on information theory, *Pattern Recognition*, <http://dx.doi.org/10.1016/j.patcog.2016.08.011>

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 galley proof before it is published in its final citable 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.

# An efficient semi-supervised representatives feature selection algorithm based on information theory

Yintong Wang<sup>a,\*</sup>, Jiandong Wang<sup>a</sup>, Hao Liao<sup>b,c,\*</sup>, Haiyan Chen<sup>a</sup>

<sup>a</sup>*College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*

<sup>b</sup>*Guangdong Province Key Laboratory of Popular High Performance Computers, College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China*

<sup>c</sup>*Department of Physics, University of Fribourg, Chemin du Musée 3, Fribourg CH-1700, Switzerland*

---

## Abstract

Feature selection (FS) plays an important role in data mining and recognition, especially regarding large scale text, images and biological data. The Markov blanket provides a complete and sound solution to the selection of optimal features in supervised feature selection, and investigates thoroughly the relevance of features relating to class and the conditional independence relationship between features. However, incomplete label information makes it particularly difficult to acquire the optimal feature subset. In this paper, we propose a novel algorithm called the Semi-supervised Representatives Feature Selection algorithm based on information theory (SRFS), which is independent of any algorithm used for classification learning, and can rapidly and effectively identify and remove non-essential information and irrelevant and redundant features. More importantly, the unlabeled data are utilized in the Markov blanket as the labeled data through the relevance gain. Our results on several benchmark datasets demonstrate that SRFS can significantly improve upon state of the art supervised and semi-supervised algorithms.

*Keywords:* Feature selection, Markov blanket, Information theory, Semi-supervised learning, Representative features

---

\*Corresponding authors

*Email addresses:* wangyintong@nuaa.edu.cn (Yintong Wang), jamesliao520@gmail.com (Hao Liao)

Download English Version:

<https://daneshyari.com/en/article/6939734>

Download Persian Version:

<https://daneshyari.com/article/6939734>

[Daneshyari.com](https://daneshyari.com)