

Author's Accepted Manuscript

Multi-criteria Feature Selection on Cost-sensitive
Data with Missing Values

Wenhao Shu, Hong Shen



PII: S0031-3203(15)00347-7
DOI: <http://dx.doi.org/10.1016/j.patcog.2015.09.016>
Reference: PR5520

To appear in: *Pattern Recognition*

Received date: 28 January 2015
Revised date: 16 June 2015
Accepted date: 7 September 2015

Cite this article as: Wenhao Shu and Hong Shen, Multi-criteria Feature Selection on Cost-sensitive Data with Missing Values, *Pattern Recognition*, <http://dx.doi.org/10.1016/j.patcog.2015.09.016>

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

Multi-criteria Feature Selection on Cost-sensitive Data with Missing Values

Wenhao Shu^d, Hong Shen^{e,f,*}

^a*School of Information Engineering, East China Jiaotong University, Nanchang, China*

^b*School of Information Science and Technology, Sun Yat-sen University, China*

^c*School of Computer Science, University of Adelaide, Australia*

Abstract

Feature selection plays an important role in pattern recognition and machine learning. Confronted with high dimensional data in many data analysis tasks, feature selection techniques are designed to find a relevant feature subset of the original features which can facilitate classification. However, in many real-world applications, missing feature values that contribute to test and misclassification costs are emerging to be an issue of increasing concern for most data sets, particularly dealing with big data. The existing feature selection approaches do not address this issue effectively. In this paper, based on rough set theory we address the problem of feature selection for cost-sensitive data with missing values. We first propose a multi-criteria evaluation function to characterize the significance of candidate features, by taking into consideration not only the power in the positive region and boundary region but also their associated costs. On this basis, we develop a forward greedy feature selection algorithm for selecting a feature subset of minimized cost that preserves the same information as the whole feature set. In addition, to improve the efficiency of this algorithm, we implement the selection of candidate features in a dwindling object set. Finally, we demonstrate the superior performance of the proposed algorithm to the existing feature selection algorithms through experimental results on different data sets .

Keywords: Feature selection, Cost-sensitive data, Multi-criteria, Incomplete data, Rough sets

*Corresponding author

Email addresses: 11112084@bjtu.edu.cn (Wenhao Shu), shenh3@mail.sysu.edu.cn (Hong Shen)

Download English Version:

<https://daneshyari.com/en/article/6940002>

Download Persian Version:

<https://daneshyari.com/article/6940002>

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