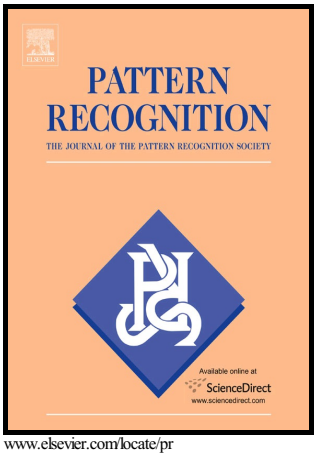
A Survey on semi-supervised feature selection methods

Razieh Sheikhpour, Mehdi Agha Sarram, Sajjad Gharaghani, Mohammd Ali Zare Chahooki



www.elsevier.com/locate/pr

PII: S0031-3203(16)30354-5
DOI: http://dx.doi.org/10.1016/j.patcog.2016.11.003
Reference: PR5946

To appear in: *Pattern Recognition*

Received date: 16 February 2016
Revised date: 3 November 2016
Accepted date: 5 November 2016

Cite this article as: Razieh Sheikhpour, Mehdi Agha Sarram, Sajjad Gharaghan and Mohammd Ali Zare Chahooki, A Survey on semi-supervised feature selection methods, *Pattern Recognition* http://dx.doi.org/10.1016/j.patcog.2016.11.003

A Survey on semi-supervised feature selection methods

Razieh Sheikhpour[1], Mehdi Agha Sarram[1*], Sajjad Gharaghani[2], Mohammd Ali Zare Chahooki[1]

[1]Department of Computer Engineering, Yazd University, Yazd, Iran

[2]Laboratory of Bioinformatics and Drug Design (LBD), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

r_sheikhpour@stu.yazd.ac.ir

mehdi.sarram@yazd.ac.ir

s.gharaghani@ut.ac.ir

chahooki@yazd.ac.ir

[*]Corresponding author: Department of Computer Engineering, Yazd University, Yazd, Iran. Tel: +989133510376

Abstract

Feature selection is a significant task in data mining and machine learning applications which eliminates irrelevant and redundant features and improves learning performance. In many real-world applications, collecting labeled data is difficult, while abundant unlabeled data are easily accessible. This motivates researchers to develop semi-supervised feature selection methods which use both labeled and unlabeled data to evaluate feature relevance. However, till-to-date, there is no comprehensive survey covering the semi-supervised feature selection methods. In this paper, semi-supervised feature selection methods are fully investigated and two taxonomies of these methods are presented based on two different perspectives which represent the hierarchical structure of semi-supervised feature selection methods. The first perspective is based on the basic taxonomy of feature selection methods and the second one is based on the taxonomy of semi-supervised learning methods. This survey can be helpful for a researcher to