Mach Learn
DOI 10.1007/s10994-013-5337-8

Unsupervised feature selection with ensemble learning

Haytham Elghazel · Alex Aussem

Received: 25 May 2012 / Accepted: 26 February 2013

Abstract In this paper, we show that the way internal estimates are used to measure variable importance in Random Forests are also applicable to feature selection in unsupervised learning. We propose a new method called Random Cluster Ensemble (RCE for short), that estimates the out-of-bag feature importance from an ensemble of partitions. Each partition is constructed using a different bootstrap sample and a random subset of the features. We provide empirical results on nineteen benchmark data sets indicating that RCE, boosted with a recursive feature elimination scheme (RFE) (Guyon and Elisseeff, Journal of Machine Learning Research, 3:1157–1182, 2003), can lead to significant improvement in terms of clustering accuracy, over several state-of-the-art supervised and unsupervised algorithms, with a very limited subset of features. The method shows promise to deal with very large domains. All results, datasets and algorithms are available on line (http://perso.univ-lyon1.fr/haytham.elghazel/RCE.zip).

Keywords Unsupervised learning · Feature selection · Ensemble methods · Random forest

1 Introduction

Feature selection is an essential component of quantitative modeling, data-driven construction of decision support models or even computer-assisted discovery. The identification of relevant subsets of random variables among thousands of potentially irrelevant and redundant variables is a challenging topic of pattern recognition research that has attracted much attention over the last few years (Hua et al. 2009; Ghaemi et al. 2009; Morais and Aussem 2010; Saeys et al. 2007; Tuv et al. 2009). In supervised learning, feature

Editors: Emmanuel Müller, Ira Assent, Stephan Günnemann, Thomas Seidl, Jennifer Dy.

H. Elghazel (✉) · A. Aussem
University of Lyon, 69622 Lyon, France
e-mail: haytham.elghazel@univ-lyon1.fr

A. Aussem
e-mail: alexandre.aussem@univ-lyon1.fr

H. Elghazel · A. Aussem
LIRIS, UMR 5205, University of Lyon 1, Villeurbanne, France

Published online: 03 April 2013