

Title: Efficient feature selection filters for high-dimensional data

Author(s): Ferreira, Artur J.^{1,3}; Figueiredo, Mário A. T.^{2,3}

Source: Pattern Recognition Letters

Volume: 33 **Issue:** 13 **Pages:** 1794-1804 **DOI:** 10.1016/j.patrec.2012.05.019

Published: Oct 1 2012

Document Type: Article

Language: English

Abstract: Feature selection is a central problem in machine learning and pattern recognition. On large datasets (in terms of dimension and/or number of instances), using search-based or wrapper techniques can be computationally prohibitive. Moreover, many filter methods based on relevance/redundancy assessment also take a prohibitively long time on high-dimensional datasets. In this paper, we propose efficient unsupervised and supervised feature selection/ranking filters for high-dimensional datasets. These methods use low-complexity relevance and redundancy criteria, applicable to supervised, semi-supervised, and unsupervised learning, being able to act as pre-processors for computationally intensive methods to focus their attention on smaller subsets of promising features. The experimental results, with up to 10(5) features, show the time efficiency of our methods, with lower generalization error than state-of-the-art techniques, while being dramatically simpler and faster. (c) 2012 Elsevier B.V. All rights reserved.

Author Keywords: Feature Selection; Filters; Dispersion Measures; Similarity Measures; High-Dimensional Data

Keywords Plus: Sparse Logistic-Regression; Feature Subset-Selection; Floating Search Methods; Multiple Data Sets; Gene Selection; Statistical Comparisons; Bound Algorithm; SVM-RFE; Classification; Redundancy

Reprint Address: Ferreira, AJ (reprint author), Inst Super Engn Lisboa, Lisbon, Portugal.

Addresses:

1. Inst Super Engn Lisboa, Lisbon, Portugal
2. Inst Super Tecn, Lisbon, Portugal
3. Inst Telecomunicacoes, Lisbon, Portugal

E-mail Address: arturj@cc.isel.ipl.pt

Funding:

Funding Agency	Grant Number
Polytechnic Institute of Lisbon	SFRH/PROTEC/67605/2010
FCT project	PEst-OE/EEI/LA0008/2011

Publisher: Elsevier Science BV

Publisher Address: Po Box 211, 1000 AE Amsterdam, Netherlands

ISSN: 0167-8655

Citation: Ferreira A J, Figueiredo M A T. Efficient feature selection filters for high-dimensional data. Pattern Recognition Letters. 2012; 13 (33): 1794-1804.