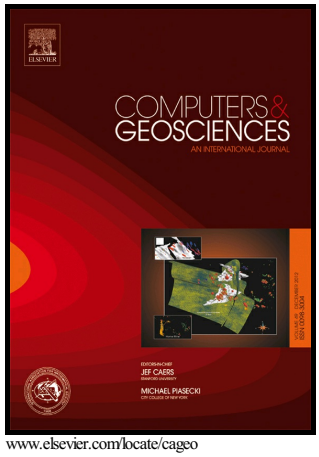
On the use of feature selection to improve the detection of sea oil spills in SAR images

David Mera, Veronica Bolon-Canedo, J.M. Cotos, Amparo Alonso-Betanzos



www.elsevier.com/locate/cageo

PII: S0098-3004(16)30862-7
DOI: http://dx.doi.org/10.1016/j.cageo.2016.12.013
Reference: CAGEO3891

To appear in: *Computers and Geosciences*

Received date: 29 March 2016
Revised date: 24 November 2016
Accepted date: 27 December 2016

Cite this article as: David Mera, Veronica Bolon-Canedo, J.M. Cotos and Amparo Alonso-Betanzos, On the use of feature selection to improve the detection of sea oil spills in SAR images, *Computers and Geosciences* http://dx.doi.org/10.1016/j.cageo.2016.12.013

On the use of feature selection to improve the detection of sea oil spills in SAR images

David Mera[a,*], Veronica Bolon-Canedo[b], J.M. Cotos[a], Amparo Alonso-Betanzos[b]

[a] *Centro Singular de Investigación en Tecnoloxías da Información (CITIUS), Universidade de Santiago de Compostela, Rúa de Jenaro de la Fuente Domínguez, 15782 - Santiago de Compostela, Spain.*
[b] *Dept. de Computación, Universidade da Coruña, Campus de Elviña s/n, A Coruña, 15071, Spain*

Abstract

Fast and effective oil spill detection systems are crucial to ensure a proper response to environmental emergencies caused by hydrocarbon pollution on the ocean's surface. Typically, these systems uncover not only oil spills, but also a high number of look-alikes. The feature extraction is a critical and computationally intensive phase where each detected dark spot is independently examined. Traditionally, detection systems use an arbitrary set of features to discriminate between oil spills and look-alikes phenomena. However, Feature Selection (FS) methods based on Machine Learning (ML) have proved to be very useful in real domains for enhancing the generalization capabilities of the classifiers, while discarding the existing irrelevant features. In this work,we present a generic and systematic approach, based on FS methods, for choosing a concise and relevant set of features to improve the oil spill detection systems. We have compared five FS methods: Correlation-based feature selection (CFS), Consistency-based filter, Information Gain, ReliefF and Recursive Feature Elimination for Support Vector Machine (SVM-RFE). They were applied on a 141-input vector composed of features from a collection of outstanding studies. Selected features were validated via a Support Vector Machine (SVM) classifier and the results were compared with previous works. Test experiments revealed that the classifier trained with the 6-input feature vector proposed by SVM-RFE achieved the best accuracy and Cohen's kappa coefficient (87.1% and 74.06% respectively). This is a smaller feature combination with similar or even