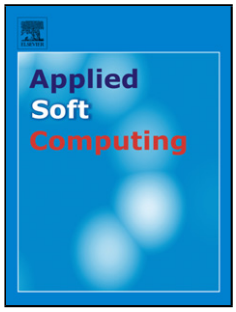

Title: Handling binary classification problems with a priority class by using support vector machines

Author: L. Gonzalez-Abril C. Angulo H. Nuñez Y. Leal

PII: S1568-4946(17)30505-7
DOI: http://dx.doi.org/doi:10.1016/j.asoc.2017.08.023
Reference: ASOC 4413

To appear in: *Applied Soft Computing*

Received date: 28-10-2016
Revised date: 29-5-2017
Accepted date: 8-8-2017

Please cite this article as: L. Gonzalez-Abril, C. Angulo, H. Nuñez, Y. Leal, Handling binary classification problems with a priority class by using support vector machines, *<![CDATA[Applied Soft Computing Journal]]>* (2017), http://dx.doi.org/10.1016/j.asoc.2017.08.023

Handling binary classification problems with a priority class by using support vector machines

L. Gonzalez-Abril[a], C. Angulo[b], H. Nuñez[c], Y. Leal[d]

[a] *Dpto. Economía Aplicada I, Universidad de Sevilla, 41018 Sevilla, Spain*
[b] *Automatic Control Dept. - Univ. Politècnica de Catalunya, 08028 Barcelona - Spain*
[c] *AI Lab. Facultad de Ciencias, Univ. Central Venezuela, 1020-A Caracas, Venezuela*
[d] *Unitat of Diabetes, Endocrinology and Nutrition (UDEN), Institut d'Investigació Biomédica de Girona (IdIBGi), 17007 Girona, Spain*

Abstract

A post-processing technique for Support Vector Machine (SVM) algorithms for binary classification problems is introduced in order to obtain adequate accuracy on a priority class (labelled as a positive class). That is, the true positive rate (or recall or sensitivity) is prioritized over the accuracy of the overall classifier. Hence, false negative (or Type I) errors receive greater consideration than false positive (Type II) errors during the construction of the model.

This post-processing technique tunes the initial bias term once a solution vector is learned by using standard SVM algorithms in two steps: First, a fixed threshold is given as a lower bound for the recall measure; second, the true negative rate (or specificity) is maximized.

Experiments, carried out on eleven standard UCI datasets, show that the modified SVM satisfies the aims for which it has been designed. Furthermore, results are comparable or better than those obtained when other state-of-the-art SVM algorithms and other usual metrics are considered.

Key words: support vector machines, post-processing strategies, pattern recognition, cost-sensitive SVM

1. Introduction

There exist situations where the correct detection of instances of one class (the positive class) is considered to be of greater importance or priority than