Multi-Objective Genetic Algorithms for Sparse Least Square Support Vector Machines

Danilo Avilar Silva and Ajalmar Rêgo Rocha Neto

Instituto Federal do Ceará, IFCE, Brasil,
Programa de Pós-Graduação em Engenharia de Telecomunicações, PPGET
daniloavilar@gmail.com, ajalmar@ifce.edu.br

Abstract. This paper introduces a new approach to building sparse least square support vector machines (LSSVM) based on multi-objective genetic algorithms (GAs) for classification tasks. LSSVM classifiers are an alternative to SVM ones due to the training process of LSSVM classifiers only requires to solve a linear equation system instead of a quadratic programming optimization problem. However, the lost of sparseness in the Lagrange multipliers vector (i.e. the solution) is a significant drawback which comes out with these classifiers. In order to overcome this lack of sparseness, we propose a multi-objective GA approach to leave a few support vectors out of the solution without affecting the classifier’s accuracy and even improving it. The main idea is to leave out outliers, non-relevant patterns or those ones which can be corrupted with noise and thus prevent classifiers to achieve higher accuracies along with a reduced set of support vectors. We point out that the resulting sparse LSSVM classifiers achieve equivalent (in some cases, superior) performances than standard full-set LSSVM classifiers over real data sets. Differently from previous works, genetic algorithms are used in this work to obtain sparseness not to find out the optimal values of the LSSVM hyper-parameters.

Keywords: Least Square Support Vector Machines, Pruning Methods, Genetic Algorithms.

1 Introduction

There are many works based on both Least Square Support Vector Machines (LSSVM) [9] and Evolutionary Computation (EC), especially Genetic Algorithms (GAs). GAs are mostly applied to optimize the LSSVM kernel parameters, as well as the classifier parameters [12,5]. Genetic Algorithms (GAs) are optimization methods and therefore they can be used to generate useful solutions to search problems. Due to the underlying features of GAs some optimization problems can be solved without the assumption of linearity, differentiability, continuity or convexity of the objective function. Unfortunately, these desired characteristics are not found in several mathematical methods when applied to the same kind of problems. GAs as a meta-heuristic method are also used to deal with classification tasks even with LSSVM.
A theoretical advantage of kernel methods such as Support Vector Machines [11] concerns the empirical and structural risk minimization which balances the complexity of the model against its success at fitting the training data, along with the production of sparse solutions [7]. Support Vector Machines (SVM) are the most popular kernel methods. The LSSVM is an alternative to the standard SVM formulation [11]. A solution for the LSSVM is achieved by solving linear KKT systems¹ in a least square sense. In fact, the solution follows directly from solving a linear equation system, instead of a QP optimization problem. On the one hand, it is in general easier and less computationally intensive to solve a linear system than a QP problem. On the other hand, the resulting solution is far from sparse, in the sense that it is common to have all training samples being used as SVs.

To handle the lack of sparseness in SVM and LSSVM solutions, several reduced set (RS) and pruning methods have been proposed, respectively. These methods comprise a bunch of techniques aiming at simplifying the internal structure of those models, while keeping the decision boundaries as similar as possible to the original ones. They are very useful in reducing the computational complexity of the original models, since they speed up the decision process by reducing the number of SVs. They are particularly important for handling large datasets, when a great number of data samples may be selected as support vectors, either by pruning less important SVs [3,4] or by constructing a smaller set of training examples [6,11], often with minimal impact on performance.

In order to combine the aforementioned advantages of LSSVM classifiers and genetic algorithms, this work aims at putting both of them to work together for reducing the number of support vectors and also achieve equivalent (in some cases, superior) performances than standard full-set LSSVM classifiers. Our proposal finds out sparse classifiers based on a multi-objective genetic algorithm which takes into account both the accuracy of classification and the support vector pruning rate. In order to do this, we also propose a new multi-objective fitness function which incorporates a cost of pruning in its formulation. Differently from previous works, genetic algorithms are used in this work to obtain sparseness not to find out the optimal values of the kernel and classifier parameters.

The remaining part of this paper is organized as follows. In Section 2 we review the fundamentals of the LSSVM classifiers. In Section 3 we briefly present some methods for obtaining sparse LSSVM classifiers. In Section 4 we introduce genetic algorithms which are necessary to understanding of our proposal in Section 5. We present our simulations and conclusions in sections 6 and 7 respectively.

2 LSSVM Classifiers

The formulation of the primal problem for the LSSVM [9] is given by

$$
\min_{\mathbf{w},\xi} \left\{ \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{\gamma}{2} \sum_{i=1}^{L} \xi_i^2 \right\},
$$

subject to \( y_i \left[ (\mathbf{w}^T \mathbf{x}_i) + b \right] = 1 - \xi_i, \ i = 1, \ldots, L \)

¹ Karush-Kuhn-Tucker systems.