

A Hyper-Solution Framework for SVM Classification: Application for Predicting Destabilizations in Chronic Heart Failure Patients

Antonio Candelieri* and Domenico Conforti*

Laboratory of Decision Engineering for Health Care Delivery, Dept. of Electronics, Informatics, Systems University of Calabria, Rende (Cosenza), Italy

Abstract: Support Vector Machines (SVMs) represent a powerful learning paradigm able to provide accurate and reliable decision functions in several application fields. In particular, they are really attractive for application in medical domain, where often a lack of knowledge exists. Kernel trick, on which SVMs are based, allows to map non-linearly separable data into potentially linearly separable one, according to the kernel function and its internal parameters value. During recent years non-parametric approaches have also been proposed for learning the most appropriate kernel, such as linear combination of basic kernels. Thus, SVMs classifiers may have several parameters to be tuned and their optimal values are usually difficult to be identified a-priori. Furthermore, combining different classifiers may reduce risk to perform errors on new unseen data. For such reasons, we present an hyper-solution framework for SVM classification, based on meta-heuristics, that searches for the most reliable hyper-classifier (SVM with a basic kernel, SVM with a combination of kernel, and ensemble of SVMs), and for its optimal configuration. We have applied the proposed framework on a critical and quite complex issue for the management of Chronic Heart Failure patient: the early detection of decompensation conditions. In fact, predicting new destabilizations in advance may reduce the burden of heart failure on the healthcare systems while improving quality of life of affected patients. Promising reliability has been obtained on 10-fold cross validation, proving our approach to be efficient and effective for an high-level analysis of clinical data.

Keywords: Heart failure, early diagnosis, classification, model selection, SVM, meta-heuristics.

1. INTRODUCTION

Support Vector Machines (SVMs) are widely adopted learning techniques, originally developed by Vapnik [1]. Basically, the approach searches for a hyper-plane able to optimally separate instances belonging to different classes while maximizing the distance among hyper-plane and instances. Optimal separating hyper-plane is expressed in terms of the closest patterns which are named Support Vectors. Accuracy/generalization trade-off performed by SVMs makes this approach rather attractive for solving classification tasks in several domains, especially in medical fields (i.e. for supporting decision making activities such as diagnosis assessment, prognosis evaluation and therapy planning). The trade-off may be handled by setting a suitable parameter, such as regularization parameter C in C-Support Vector Classification (C-SVC).

In order to obtain non-linear separation, SVMs exploit the kernel trick, that transforms non-linearly separable problems into high dimensional possibly linearly separable ones. Parameters tuning (regularization, kernel type and its internal parameters) is a crucial issue for SVMs learning task, but choice is usually quite difficult to be optimally identified a-priori. Generally, a set of suitable values is fixed for each parameter, providing a “grid” of possible configurations by which performing several implementations

of the learning task. However, grid approach may be really time consuming, because also bad-performing configurations have to be tested.

In recent years, meta-heuristics techniques (such as Tabu-Search and Genetic Algorithm) have been proposed in order to improve the best-tuning process for SVMs [2,3]. At the same time, also non-parametric approaches have been suggested for learning kernel function as combination of a preselected set of basic kernels (i.e. linear or convex combinations) using Semi Definite Programming (SDP) [4,5] or Genetic Algorithms [6]. The common idea of the parametric and non-parametric approaches (also known as Model Selection and Multiple Kernel Learning, respectively) is to specialize a single SVM classifier, identifying the most reliable parameters setting up for a SVM with fixed kernel, or the most suitable kernel function obtained as combination of a set of kernels.

An alternative and quite promising strategy is to combine different good-performing classifiers in order to exploit the advantages offered by each of them and reduce the risk to make classification errors on new unseen data [7-9]. Concerning this task (also known as Ensemble Learning), meta-heuristics based approaches have been recently proposed for optimizing a weighted voting systems [10] of SVM classifiers [11].

In this paper, we present an original “hyper-solution” SVMs classification framework able to perform, at the same time, Model Selection, Multiple Kernel Learning and Ensemble Learning. The idea is to take advantages of Tabu-Search (TS) or Genetic Algorithm (GA) meta-heuristics, in

*Address correspondence to these authors at the Laboratory of Decision Engineering for Health Care Delivery, Dept. of Electronics, Informatics, Systems University of Calabria, Rende (Cosenza), Italy; Tel: +39-984-494732; E-mails: acandelieri@deis.unical.it, mimmo.conforti@unical.it

order to more efficiently search for the best hyper-SVM classifier, avoiding at the same time to test bad-performing configurations. The best decision model provided by the framework can be a SVM with a base kernel, a SVM with a linear combination of base kernels, or an ensemble of different base SVM classifiers. This basically motivates the use of the term “hyper-solution”.

The proposed framework has been specifically developed and used in order to identify an objective and simple criteria for predicting new destabilizations in Chronic Heart Failure (CHF) patients: a crucial mid-long term goal for the management of these patients. Under this respect we have used a dataset of 301 instances. Each one is a collection of some clinical parameter measurements, easy to be acquired, in an automatic or semi-automatic way, also directly by patient at home. For each visit (instance), patient was in a stable condition and the class label was identified by the condition at two weeks later evaluated by clinical domain experts.

The paper is organized as follows. In section 2, we briefly outline the main characteristics of chronic heart failure clinical domain. In section 3, we describe the relevant data set, the implemented meta-heuristics (Tabu-Search and Genetic Algorithm) and the experimental framework. The obtained results and relevant discussion is the content of section 4. Finally, some conclusions complete the paper.

2. CHRONIC HEART FAILURE CLINICAL DOMAIN

Chronic Heart Failure (CHF) is a complex clinical syndrome which impairs the ability of the heart to pump sufficient blood to cover the body’s metabolic needs. This can severely affect patient’s ability in their normal daily activities, and it is a leading cause of hospital admissions and deaths, as well as a strong impact in terms of social and economic effects [12,13].

CHF prevalence increases rapidly with age [14], with a mean age of 74 years, and the increase in the proportion of elderly population concurs in rising prevalence of heart failure. In particular, according to the European Society of Cardiology, there are at least 10 millions of patients with heart failure in the European countries.

Even though heart failure is a chronic syndrome, it does not evolve gradually. Periods of relative stability alternate with acute destabilizations. During a stable phase the crucial mid-long term goal should be to predict and, hopefully prevent, destabilizations and death of the CHF patient, reducing the burden of heart failure on the healthcare system while improving quality of life of the affected patients.

Although a system of frequent monitoring could be useful for clinicians to recurrently evaluate patient conditions and eventually provide intervention before CHF patient becomes as severely ill as require re-hospitalization [15], symptoms and signs remain often difficult to identify. Moreover, only by applying sophisticated methods is possible to predict, define and assess new decompensation events, while the general interest is to find simple methods and criteria on how to predict when patient will further decompensate.

For such reasons, machine learning techniques may be a practical and effective solution for generating and proving new hypothesis, mining and generalizing new medical knowledge directly from pertinent real examples.

Under this respect, we developed and tested a knowledge discovery task for the extraction of new decision models able to early detect new decompensation events [16,17]. The knowledge discovery task was specifically devised and performed within the activities of the EC FP6 funded project HEARTFAID (www.heartfaid.org), aiming at developing an innovative knowledge based platform for more effective and efficient clinical management of heart failure within elderly population. Decision models extracted through knowledge discovery task were embedded into the Clinical Decision Support System (CDSS) of HEARTFAID, integrating the platform’s knowledge base. We implemented different machine learning approaches, such as Decision Trees, Decision Lists, Support Vector Machines and Radial Basis Function Networks.

In this paper, we further extend our research activity, by proposing the SVMs hyper-solution framework as effective tool able to early predict CHF patient destabilizations.

3. MATERIALS AND METHODS

3.1. Data Set Description

The considered data set is related to 50 patients frequently visited (every two weeks) and clinically assessed by the medical doctor. During each visits, relevant clinical parameters (Systolic Blood Pressure, Heart Rate, Respiratory Rate, Body Weight, Body Temperature, Total Body Water) are acquired and stored, while the patient health conditions, with respect to stable or decompensated status, are evaluated and assessed.

After a data cleansing step, we have obtained a dataset with 301 instances (visits for stable patients) where class labels are related to the patient condition at two weeks. Hence, classification problem has been defined as searching for a relationship between clinical parameters value of a stable CHF patient and the risk of a further decompensation in two weeks. Further general information about patients has been also taken into account for the analysis (gender, age, smoking and alcohol use).

3.2. Tabu-Search

Tabu-Search (TS) [18] is an iterative neighbourhood search method for solving optimization problems. The general step of the technique consists in searching, from a current solution, the next best one in a given neighbourhood. In order to avoid local minimum trapping, the new solution may be less efficient than the previous one. However, this strategy may select recently visited solutions, showing cycles. Under this respect, TS uses a short memory (*tabu list*) to avoid such a kind of undesired behavior, avoiding to select the most recently visited solutions (*tabu solutions*). In a promising region of the space, TS allows extensive search to find a best solution (*intensification* phase), elsewhere enables large changes of the solution in order to find quickly another promising region (*diversification*). These two strategies are usually applied alternatively.

3.3. Genetic Algorithm

Genetic Algorithm (GA) [19] emulates the natural evolution process into a population. As the most “fit” individuals have high probability to transfer their features to the next generation (*offspring*), so the most “good” solutions have more chances to “generate” new good ones, hopefully the best one (optimum) after several generations. Each solution is defined by a string of values (generally bits) and is named *chromosome*. Each value of the string is a *gene*. Generating new individuals is performed by selecting two of the most fit ones into the current population (parents), *crossing-over* them for producing two new solutions and finally applying a random *mutation* on some of their genes. Size of population, selection strategy, crossing-over, mutation and termination criteria regulate the performance of the meta-heuristic.

In this paper we developed a new implementation of the canonical GA formulation. In particular, we have not adopted bits but numeric representation for genes, in two different settings: *nominal* genes, where each gene can assume only values into a prefixed set; and *mixed* genes, where some genes are not nominal but numeric and can assume any real continuous value within a prefixed range. This aspect is really useful for instantiating SVM parameters, such as regularization, γ parameter for RBF kernel, coefficients for kernel combination and weights for the SVMs ensemble.

3.4. Experimental Framework

In our experiments, for framework based on TS and GA with nominal genes, we have adopted:

- $C = \{1, 2, 5, 10\}$;
- Kernel = {RBF, Normalized Polynomial, Linear};
- Degree (Normalized Polynomial Kernel) = {2, 3, 4, 5};
- γ (RBF kernel) = {0.001, 0.002, 0.005, 0.007, 0.01, 0.025, 0.05, 0.075, 0.1, 0.25, 0.5, 0.75, 1, 2.5, 5, 7.5, 10, 25, 50, 75, 100, 250, 500, 750, 1000}
- coefficients (Multiple Kernel Learning) and weights (Ensemble Learning) = {0.1, 0.25, 0.5, 0.75, 1};

For framework based on GA with mixed genes only some parameters may be defined as numeric: C , γ , coefficients and weights, for which the selected ranges has been maintained. Then, we combined three different kernels for Multiple Kernel Learning and three SVMs for Ensemble Learning.

It is worth while to point out that classes are strongly unbalanced in our dataset, as there are only 11 instances labelled as “decompensation in 2 weeks” and 290 as “stability will be maintained in 2 weeks”, but predicting correctly the minority class is the most relevant and critical goal. For such a reason, we have adopted also a cost sensitive classification approach, providing higher misclassification costs for instances belonging to the minority class.

Taking advantages from the proposed solution, we have stated “cost” as a further parameter to be optimized, in particular we have established cost = {14, 16, 18, 22} for framework based on TS and GA with nominal genes, while

cost is any continuous real value between 14 and 22 for framework based on GA with mixed genes. Cost’s range has been identified by previous analysis performed on the data [16,17]. Unbalance in classes distribution has been also taken into account in evaluating solution reliability, by using the Balanced Classification Accuracy (BCA). Such a measure is the average of the accuracies on each class, and assigns higher penalization to classifications errors related to instance belonging to the minority class.

As far as the meta-heuristic parameter selection is concerned, the following settings have been used:

Tabu-Search

- Tabu-list length = 10;
- 5 consecutive intensification iterations with no improvements for starting the diversification;
- 5 consecutive diversification iterations with no improvements for stopping intensification-diversification iteration;
- stopping criteria: BCA (solution benefit)=100% or 30 intensification-diversification iterations.

Genetic Algorithm with Nominal Genes

- population size = 10 chromosomes;
- (strong) mutation probability = 0.001;
- stopping criteria: BCA (fitness)=100% or 10 consecutive iterations with no improvements or population homogeneity or 100 total iterations.

For Genetic Algorithm with Mixed Genes, we have made exploration more powerful, since searching space is larger with respect to Genetic Algorithm with nominal genes:

- populations size = 30 chromosomes;
- (strong) mutation probability = 0.01;
- stop criteria: the same of the case with nominal genes with 30 consecutive iterations without improvements.

4. RESULTS AND DISCUSSION

In this section we report the results obtained by the application of our proposed framework on the available data set. In Table 1, some performance indexes evaluated on a stratified 10 fold-cross validation (BCA, Sensitivity and False Positive Rate) are reported.

First of all, both of two GA-based frameworks have proved to be more effective and reliable than TS-based one, identifying a solution with higher BCA, which is the optimization objective for meta-heuristics (fitness for GA and benefit for TS). Moreover, GA-based frameworks have also proved to be more sensible (more accurate on minority class) than TS.

GA-based framework with nominal genes has showed to be the most effective (highest BCA) and most efficient approach, consuming about 1/2 of the learning time spent by TS and GA with mixed genes.

Learning models provided by the framework are: an SVMs ensemble for GA with nominal genes (best model), an SVM with Multiple Kernel Learning for TS and an SVM

Table 1. Performance Indexes for the Three Framework Implementations

| | Balanced Classification Accuracy* | Sensitivity | False Positive Rate | Time |
|---|-----------------------------------|---------------|---------------------|----------------|
| TS-based framework | 85.56% | 81.82% | 10.69% | 2:51:44 |
| GA-based SVM framework (with nominal genes) | 87.35% | 90.91% | 16.21% | 1:42:40 |
| GA-based SVM framework (with mixed genes) | 86.49% | 90.91% | 17.93% | 3:02:39 |

*Performance measure for optimization (fitness and benefit of solutions, for GA e TS respectively).

with base kernel obtained by Model Selection for GA with mixed genes.

A relevant result of our experiments is related to the framework based on GA with mixed genes: the obtained model (a base SVM) is more reliable than one obtained by a “grid-approach” Model Selection analysis previously performed on the same dataset (BCA=67.62%, Sensitivity=54.55% and False Positive Rate=19.31%, on 10-fold cross validation), proving that “continuous” genes may allow to identify more reliable solutions.

Finally, results suggest that proposed framework may drastically reduce computational time for analysing several classifier configurations in multiple different learning tasks (Model Selection, Multiple Kernel Learning and Ensemble Learning), assuring that a “sub-optimal” solution may be achieved even though searching for in very large solution spaces.

Solution by TS-Based Framework

SVM classifier through Multiple Kernel Learning:

- Cost for misclassifications on minority class = 22.0;
- Regularization C=10.0;
- Multiple Kernel Learned = Normalized Polynomial Kernel with degree value 2 + 0.5xNormalized Polynomial Kernel with degree value 3 + Normalized Polynomial Kernel with degree value 3.

Solution by GA (with Nominal Genes)-Based Framework

SVM classifiers Ensemble:

- Costs for misclassifications on minority class for each SVM classifier = 20.0, 14.0, 14.0;
- Regularization C for each SVM classifier: 5.0, 10.0, 2.0;
- Kernel ensemble = 0.5xNormalized Polynomial Kernel with degree value 4 + 0.5xNormalized Polynomial Kernel with degree value 4 + 0.5xRBF Kernel with gamma value 750.

Solution by GA (with Mixed Genes)-Based Framework

Base SVM classifier:

- Cost for misclassifications on minority class = 22.0;
- Regularization C = 6.4501;
- Normalized Polynomial Kernel with degree value 5.

It is worthwhile to observe that the best misclassification costs selected by the three framework implementations are always quite close to the ratio between the classes size (290/11 = 26.36).

5. CONCLUSIONS

We have presented a hyper-solution framework for SVM classification learning, based on meta-heuristics, and able to perform, at the same time, Model Selection, Multiple Kernel Learning and Ensemble Learning with the aim to identify the best hyper-classifier.

We have adopted Tabu-Search and Genetic Algorithm for optimizing the best solution search process. We have also proposed some modifications to the canonical formulation of GA, in order to better handle SVM’s specificities.

We have compared our different implementations of the framework, applying them on a dataset (301 instances) related to a crucial medical problem: the early detection of decompensation conditions in Chronic Heart Failure patients.

We have obtained, in reasonable time, a reliable hyper-classifier by each framework implementation. The best one shows high Balanced Classification Accuracy (87.35%) and Sensitivity (90.91%), evaluated on stratified 10 fold cross validation. Such a decision model (an optimized ensemble of SVMs) may be potentially useful for predicting risk of a new destabilization in 2 weeks for CHF patients, supporting clinical experts in recognizing worsening in advance and preventing re-hospitalization.

REFERENCES

- [1] Vapnik V. Statistical Learning Theory. New York, Wiley, 1998.
- [2] Huang CL, Wang CJ. A GA-based feature selection and parameters optimization for support vector machines. *Expert Syst Appl* 2006; 31:231-40.
- [3] Lebrun G, Charrier C, Lezoray O, Cardot H. Tabu search model selection for SVM. *Int J Neural Syst* 2008; 18:19-31.
- [4] Lanckriet G, Cristianini N, Bartlett P, El Ghaoui L, Jordan MI. Learning the kernel matrix with semidefinite programming. *J Mach Learn Res* 2004; 5:27-72.
- [5] Conforti D, Guido R. Kernel based support vector machine via semidefinite programming: Application to medical diagnosis. *Comput Operat Res* 2010; 37(8): 1389-94
- [6] Lessmann S, Stahlbock R, Crone SF. Genetic algorithms for support vector machine model selection. *International Joint Conference on Neural Networks* 2006.
- [7] Dietterich TG. Ensemble methods in machine learning. In: Kittler J, Roli F, Eds. *Multiple Classifier Systems. First International Workshop, MCS 2000, Cagliari, Italy, Springer-Verlag* 2000; pp. 1-15.
- [8] Condorcet NC. *Essai sur l’application de l’analyse a la probabilité des decisions rendues a la pluralité des voix*. Imprimerie Royale 1785.
- [9] Valentini G, Masulli F. Ensembles methods. *Lect Notes Comput Science* 2002; 2486: 3-20.
- [10] Lam L, Sue C. Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Trans Syst Man Cybernetics* 1997; 27(5): 553-68.
- [11] He LM, Yang XB, Kong FS. Support vector machines ensemble with optimizing weights by genetic algorithm. *Proceedings of the*

- Fifth International Conference on Machine Learning and Cybernetics, 2006.
- [12] Hunt SA, Abraham WT, Chin MH, *et al.* ACC/AHA 2005 Guideline Update for the Diagnosis and Management of Chronic Heart Failure in the Adult: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* 2005; 112(12): e154-235
- [13] Swedberg K, Cleland J, Dargie H, *et al.* Guidelines for the diagnosis and treatment of chronic heart failure: executive summary (update 2005): The Task Force for the Diagnosis and Treatment of Chronic Heart Failure of the European Society of Cardiology. *Eur Heart J* 2005; 26(11): 1115-40.
- [14] McKee PA, Castelli WP, McNamara PM, *et al.* The natural history of congestive heart failure: the Framingham study. *N Engl J Med* 1971; 285:1441-6.
- [15] Chaudhry SI, Barton B, Mattera J, Spertus J, Krumholz HM. Randomized trial of Telemonitoring to Improve Heart Failure Outcomes (Tele-HF): study design. *J Card Fail* 2007; 13(9): 709-14.
- [16] Candelieri A, Conforti D, Perticone F, Sciacqua A, Kawecka-Jaszcz K, Styczkiewicz. Early detection of decompensation conditions in heart failure patients by knowledge discovery: the heartfaid approaches. *Comput Cardiol* 2008; 35: 893-6.
- [17] Candelieri A, Conforti D, Sciacqua A, Perticone F. Knowledge discovery approaches for early detection of decompensation conditions in heart failure patients. 9th Int. Conf. on Intelligent Systems Design and Applications, 30/11 – 2/12, 2009, Pisa (IT), DOI: 10.1109/ISDA.2009.204, 357-362, 2009.
- [18] Glover F. Future paths for integer programming and links to artificial intelligence. *Comput Operat Res* 1986; 13: 533-49.
- [19] Holland JH. Adaption in natural and artificial systems. Ann Arbor: University of Michigan Press 1975.

Received: December 5, 2009

Revised: March 5, 2010

Accepted: March 15, 2010

© Candelieri and Conforti; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.