Towards a Surrogate-Assisted Multi-Objective Full Model Selection

Alejandro Rosales-Pérez[1], Jesus A. Gonzalez[1], Carlos A. Reyes-Garcia[1], and Carlos A. Coello Coello[2]

[1] Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE)
Computer Science Department
Luis Enrique Erro No 1, Sta. Ma. Tonantzintla, C.P. 72840, Puebla Pue., Mexico.
{arosales,jagonzalez}@inaoep.mx

[2] Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV-IPN)
Computer Science Department
San Pedro Zacatenco, Mexico City, Mexico
ccoello@cs.cinvestav.mx

Abstract. This research approaches the full model selection problem. The full model selection problem is defined as a method in which, given a pool of pre-processing methods, feature selection and learning algorithms, to choose from, we select a combination of them, together with their hyper-parameters,in such a way, that we can provide the "best" generalization performance on a given dataset. We propose to face this as a multi-objective optimization problem, where the classification-error and the model complexity are defined as the objectives to be minimized. We propose to use a surrogate-assisted multi-objective evolutionary algorithm approach to explore the models space. Our proposal derives from the fact that estimating the values of the objective function could be computationally expensive. Therefore, by using surrogate-assisted optimization we expect to reduce the number of full models that should be trained and tested so that we can reduce the total number of fitness function evaluations, without degrading, in a significant manner, the quality of the models. Our preliminary results give evidence of the validity of our proposed approach.

Key words: Full model selection, multi-objective optimization, ensemble methods, VC dimension.

1 Introduction

Classification is a mainstream in supervised learning. A large number of learning algorithms have been proposed so far, with the aim of constructing a classification model. However, there does not exist a single learning algorithm that is the best for all problems; this is sometimes known as the **no free lunch theorem** [23]. In addition to this lack of a universal best algorithm, the performance of many of them highly depends on the fine-tuning of a set of hyper-parameters. This raises the issue of model selection.

It is noteworthy that besides the learning algorithm, there exist methods for pre-processing the data and for feature selection, which could help to improve the model performance. For instance, k-nearest neighbor method is not robust to the way the features are scaled. Therefore, it could result beneficial if the data are first normalized or standardized. Moreover, it could also be favorable if the irrelevant/redundant features are previously filtered. Based on the above, one faces the issue of choosing a combination of these methods together with the hyper-parameters that improve the performance of the model. This is a problem known as *full model selection* [11].

In the literature, there are several studies on the model selection problem. Some of these have approached it as an optimization problem. They could be differentiated in two essential aspects: the criterion used and the search engine adopted for this task. Regarding the first aspect, this problem has been tackled both as a single criterion optimization problem and as a multiple criteria optimization problem. The single criterion approaches typically utilize the well-known k-fold cross validation to estimate the model performance [1,3–5,11]. On the other hand, multiple criteria approaches consider an estimation of the model performance and a measure of its complexity [2, 20]. Others have considered to minimize the error rates on positive and negative classes [6, 14], estimates of the bias-variance model [18,19], or different estimates of the model performance [13].

Concerning the second aspect, authors have investigated the use of grid search [4, 21], gradient-based methods [1, 5], and bio-inspired meta-heuristics such as evolutionary algorithms [6,13,14,18–20], artificial immune systems [2], or particle swarm optimizers [3, 11]. Grid search is the simplest approach to adjust the hyper-parameters values. Under this strategy, each combination of hyper-parameters is tested, which makes this approach suitable to adjust only a few number of hyper-parameters. In full model selection, several hyper-parameters need to be adjusted simultaneously, which could be unsuitable for this approach.

In spite of the fact that gradient-based methods are more efficient and they have been successfully applied to model selection problems, they still have several shortcomings. For instance, the objective function must be differentiable with respect to the hyper-parameters. Furthermore, the effectiveness of these kinds of methods highly depends on the initial search point. This makes that these methods are susceptible to getting trapped in a local optimal solution.

Evolutionary algorithms have also been used in previous studies for model selection. These kinds of algorithms could be less susceptible to local optimal solutions than gradient-based methods. Although they could be cheaper than grid search methods, their computational cost could still be high.

An alternative approach formulates the model selection problem as a supervised learning one by constructing a meta-model, which is in charge of making the suggestion for models. Recent studies have combined the ideas of treating model selection as supervised learning and optimization problems [12,15,17]. The main idea under these hybrid approaches is to use the meta-model for obtaining suggestions of potential models to be used as initial points in the search step. However, the quality of the meta-model depends on the quality of the samples as well as on the number of problems which are learned and could be limited. These shortcomings could affect the convergence in the optimization step.

In spite of the considerable number of studies on model selection, most of them have focused on single model selection (i.e., the learning algorithm is fixed a priori and the task is performing the selection of its hyper-parameters), which could not be the most suitable for a particular problem. The studies on full model selection are still scarce, and they have been formulated as a single criterion optimization problem. Nevertheless, the advantages of multiple criteria over a single criterion on hyper-parameters tuning have been pointed out by several authors [6, 13].

Inspired by the above, in this research we propose to tackle both the full model selection problem as a multi-objective one (i.e., to consider multiple criteria) and the computational cost in this task. The latter is addressed by using surrogate-assisted optimization. The main motivation of this research is, precisely, to design an algorithm to perform a multi-objective full model selection emphasizing its efficiency, measured in terms of number of evaluations performed. Our working hypothesis is that, by minimizing simultaneously the error and complexity of a full model through surrogate-assisted optimization, it will be possible to obtain, in an efficient way, accurate full models that satisfy a good trade-off between the considered criteria. The estimation of the complexity should be generic in order to make it feasible to the full model selection problem, which is one of the main challenges in this research. The main contribution of this research is a general model selection framework, whose formulation makes it applicable to any learning algorithm and, in consequence, to the full model selection problem. Additional contributions are the following: (i) the multi-objective formulation of the full model selection problem (i.e., to choose a combination of pre-processing, feature selection methods, and learning algorithm together with its hyper-parameters); (ii) the hybridization with surrogate-assisted optimization to reduce the number of objective functions evaluations; and (iii) since the outcome of the multi-objective optimization is a set of solutions that satisfy a good trade-off between the objectives, the strategies to address the final model construction from such set would also be an additional contribution.

The remainder of this paper is as follows: Section 2 describes the basic concepts related to evolutionary multi-objective optimization. Section 3 describes the proposed research methodology. Section 4 shows the preliminary results of our research to give evidence of the feasibility of our proposal. Finally, Section 5 details some conclusions and indicate paths of future research.

2 Evolutionary Multi-Objective Optimization

A multi-objective optimization problem (MOP) is stated as follows [7]:

$$\begin{aligned} \text{minimize } & \mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), \ldots, f_l(\mathbf{x})]^T \\ \text{subject to } & \mathbf{x} \in \mathcal{X} \end{aligned} \tag{1}$$

where $\mathbf{x} = [x_1, \ldots, x_n] \in \mathbb{R}^n$ is a vector of decision variables, $f_i(\mathbf{x})$, $i = 1, \ldots, l$, are the l-objective functions, and $\mathcal{X}$ is the set of feasible solutions.

In a MOP, the objectives could be in conflict. In such cases, the notion of optimum refers to finding good trade-offs among the objectives. The most accepted notion of optimality is the one proposed by Pareto. To describe the concept of Pareto optimality, we will introduce the following definitions:

Definition 1. Pareto dominance: *A solution* $\mathbf{x^{(1)}}$ **dominates** *a solution* $\mathbf{x^{(2)}}$ *(denoted by* $\mathbf{x^{(1)}} \preceq \mathbf{x^{(2)}}$*) iff* $\mathbf{x^{(1)}}$ *is better than* $\mathbf{x^{(2)}}$ *at least in one objective and it is not worse in the rest.*

Definition 2. Pareto optimality: *A solution* $\mathbf{x}^* \in \mathcal{X}$ *is a* **Pareto Optimal** *if there does not exist another solution* $\mathbf{x}' \in \mathcal{X}$ *such that* $\mathbf{x}' \preceq \mathbf{x}^*$*.*

The Pareto optimal definition does not produce a single solution, but a set of them, which represent the possible trade-offs among the different objectives. The set of trade-off solutions (in decision variable space) is known as **Pareto optimal set**.

Definition 3. Pareto optimal set: *The Pareto optimal set (PS) is defined as:*

$$PS = \{\mathbf{x} \in \mathcal{X} \mid \mathbf{x} \textit{ is a Pareto optimal solution}\}$$

The objective function values corresponding to the elements of the Pareto optimal set constitute the so-called **Pareto front**. Formally,

Definition 4. Pareto front: *The Pareto front (PF) is defined as:*

$$PF = \{\mathbf{f}(\mathbf{x}) \mid \mathbf{x} \in PS\}$$

Evolutionary algorithms have gained popularity to solve MOPs, mainly because they can obtain several elements of the Pareto optimal set in a single run. Furthermore, they are less sensitive to the shape and continuity of the Pareto front than mathematical programming techniques. In the literature, a large number of multi-objective evolutionary algorithms (MOEAs) have been reported so far. NSGA-II [9], PESA-II [8], and MOEA/D [24] are some of these MOEAs. A comprehensible review of MOEAs can be found in [7].

In the full model selection problem, both pre-processing, feature selection methods, and the learning algorithm together with its hyper-parameters have to be chosen, resulting in a vast search space. Furthermore, two criteria should be simultaneously optimized (the model performance and the model complexity). Thus, stochastic search techniques, such as MOEAs, are well suited for this. In spite of the MOEAs' advantages, they have to perform a relatively high number of fitness function evaluations to get a reasonable approximation to the Pareto front. This could be a shortcoming in the problem that we face, since the computation of the objective could require to train and to test a model a number of times. To overcome this handicap, in this research we propose to study the surrogate-assisted multi-objective evolutionary optimization to address this issue. A surrogate is a cheaper approximation to the fitness function and it is used to approximate the fitness value of a given model. By using surrogate-assisted optimization, we expect to reduce the number of solutions evaluated with the fitness function and, in this manner, to reduce the computational cost of this task. Next, we explain the proposed research methodology.

3 Research Methodology

1. **Design an algorithm for multi-objective full model selection.** This stage of the research involves a review of the literature in order to find how to estimate the model complexity in a general fashion to any learning algorithm. This stage involves also analyzing the advantages and disadvantages of each approach and choosing one according to the previous analysis. It also involves the formulation of the full model selection problem as a multi-objective one, which implies the definition of how solutions are represented into the MOEA, the operators adopted to evolve the models, and the strategy for exploring the models space. The integration of these in an algorithm and its evaluation are also tasks in this stage. An analysis of the performance is used to propose improvements to the algorithm.
2. **Design a strategy for decision making in multi-objective full model selection.** This stage is mainly focused on analyzing the non-dominated front in order to determine what solutions should be chosen as the final classification model. We propose to explore two alternatives: the first one consists on choosing a single model from those generated during the optimization step. The second one considers an ensemble of models. In the first one, it is necessary to identify the regions on the Pareto front so as to find in which region is located the model with the best performance on unseen data. On the other hand, the second approach involves to study strategies to choose the subset of accurate and diverse models to be used in the ensemble. Both approaches are studied in this stage, and their advantages and disadvantages are also analyzed. The improvements are based on the results of the performed analysis. In case of being necessary, modifications to stage one are also performed.
3. **Integration with a surrogate-assisted optimization approach.** This stage includes the hybridization of the MOEA with a surrogate, which is used to approximate the fitness values of the models. Strategies to make such hybridization and an interaction with the expensive fitness functions are proposed in this stage. The integration of the proposed scheme with the multi-objective full model selection approach is also considered in this stage. We evaluate the performance of the proposed algorithm in terms of its accuracy-performance and the number of fitness function evaluated. Improvements to this stage are based on the performed evaluation. If necessary, modifications to the previous stages are also considered.

By following this methodology, we expect to achieve in a successful manner the goals of this research. The next section presents the preliminary results reached to date.

4 Results Achieved

In this section, we describe the preliminary results of our research. First, we present a brief description of the proposed method to deal with the model selection problem. Next, we present experimental results together with a statistical analysis.

4.1 Towards a Multi-Objective Full Model Selection

Following the proposed research methodology, we have formulated the model selection problem as a multi-objective optimization one. We consider different kinds of learning algorithms together with their hyper-parameters. For doing so, we first need to estimate the model complexity in a general fashion to any learning algorithm. We studied two approaches to do this. The first one is the model variance, due to the fact that a high complex model has a high variance. The second one is the VC-dimension, a measure of the capacity of the model, which is also related to the model complexity. The studies related to the variance as a measure of the model complexity are reported in [18,19]. Regarding the VC-dimension as a measure of the model complexity, we have proposed an approach for multi-objective model type selection (i.e., both a learning algorithm and its hyper-parameters are chosen). We compared three MOEAs widely used in the literature. These MOEAs are NSGA-II [9], PESA-II [8], and MOEA/D [24]. In the comparison, these algorithms reached, on average, a very similar performance in the problem at hand. However, the computational cost of MOEA/D was lower than the others. For this reason, we adopted MOEA/D.

In evolutionary algorithms, the solutions must be encoded in individuals. We propose to encode the solutions in a D-dimensional vector, where $D = 7$, as follows:

$$\mathbf{x}^i = \left[x_m^i, x_{hp_1}^i, \ldots, x_{hp_{D-1}}^i\right] \tag{2}$$

where x_m^i controls the learning algorithm, and $[x_{hp_1}^i, \ldots, x_{hp_{D-1}}^i]$ represents the hyper-parameters for the learning algorithm.

The fitness function that we propose to estimate the merit of each model for a given dataset is as follows:

$$\begin{aligned} err &= \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}\left(y_i, y_i^*\right) \\ complexity &= \operatorname*{argmin}_h \sum_{i=1}^{k} \left[\xi\left(n_i\right) - \Phi\left(n_i/h\right)\right]^2 \end{aligned} \tag{3}$$

where N is the number of samples in the training set, y_i is the class label, y_i^* is the class predicted by the model, $\mathcal{L}(y_i, y_i^*)$ is a loss function, $\xi(n_i)$ is the experimental maximum deviation error rate of two observed independent labeled data sets, and $\Phi(n_i/h)$ is the expectation of the largest deviation error between two sets (we refer to [22] for details about complexity estimation). We used the 0/1 loss function because it is well suited for classification tasks.

These definitions correspond to the first step of our research methodology. Considering the second step, we have proposed three strategies for constructing a final classification model from those that are in the non-dominated set. The first strategy consists in choosing a single solution from the non-dominated front. The second and third strategies are based on the idea of combining the multiple models in the non-dominated front in an ensemble. For the first strategy, we analyzed the performance on test sets of each solution in the non-dominated front. We empirically noted that the solutions that are in the knee of the curve have the best generalization. We also noted that this solution in most cases corresponds to the one closest to the (0,0) point. Therefore, the objectives are first normalized and then the Euclidean distance is computed between each point and the (0,0) point. The one with the minimum distance is chosen. In Figure 1, the solution that was selected with this strategy is circled.

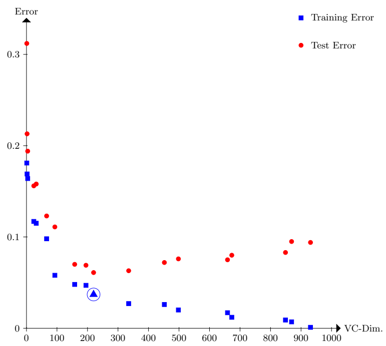

Fig. 1: Behavior of non-dominated solutions on training samples and test samples

The second strategy consists in considering all solutions in the non-dominated front and combining them in an ensemble. The final output of the ensemble is the weighted linear aggregation of the individuals predictions given by each model. The weight of each model is assigned based on the distance from such model to the (0,0) point, in objective function space. The third strategy considers to choose a subset of models in the non-dominated set taking into consideration the diversity among them. Next, we present the experimental results with our approach and the comparative study between these three strategies.

4.2 Experimental Results

We performed experiments using the IDA benchmark repository datasets. This benchmark has 13 datasets of binary classification problems. Table 1 shows some characteristics of these datasets. These datasets were previously pre-processed by [16], in which each data set was divided in 100 partitions for training and test (20 for the cases of image and splice data sets). We performed the model selection independently for each replication of each dataset.

Table 1: Details of the data sets used in our experiments.

ID	Data set	Feat.	Training Samples	Testing Samples	Replications
1	Banana	2	400	4900	100
2	Breast Cancer	9	200	77	100
3	Diabetes	8	468	300	100
4	Flare Solar	9	666	400	100
5	German	20	700	300	100
6	Heart	13	170	100	100
7	Image	20	1300	1010	20
8	Ringnorm	20	400	7000	100
9	Splice	60	1000	2175	20
10	Thyroid	5	140	75	100
11	Titanic	3	150	2051	100
12	Twonorm	20	400	7000	100
13	Waveform	21	400	4600	100

The performance of the proposed model selection method is assessed by means of the error rate attained on each data set. We compare the three strategies for the final classification model construction, and the best one is compared with PSMS, a full model selection method reported in the literature.

Table 2 shows the average error rates and standard error reached by our three strategies: single model selection (MOMTS-S1), ensemble of the whole non-dominated front (MOMTS-S2), and the ensemble of some solutions in the non-dominated front taking into consideration the diversity (MOMTS-S3). It also shows the performance reached by PSMS. From this table, one could note that the best results among the three strategies is reached by MOMTS-S2, the ensemble approach that combines all solutions in the non-dominated front. This is not entirely surprising, since the benefits of using the ensemble method for improving model performance are well known. For assessing the statistical difference between the three approaches for the final model construction over the different data sets, Demšar [10] recommends Friedman's test for comparing multiple classifiers over multiple data sets. This test is performed with a 95% of confidence, and the Nemenyi test as the post hoc test. According to these tests, the ensemble of the whole front approach is found to be statistically superior to the others.

Comparing with PSMS, we noted that MOMTS-S2 gets better performance in 12 out of 13 data sets. This shows the advantages of using a multi-objective

Table 2: Results obtained by the proposed approach, and those obtained by PSMS. The best result for each data set is shown in **boldface**.

ID	PSMS [11]	MOMTS-S1	MOMTS-S2	MOMTS-S3
1	11.08 ± 0.083	14.34 ± 0.105	$\mathbf{10.48 \pm 0.046}$	12.91 ± 0.160
2	33.01 ± 0.658	29.89 ± 0.736	$\mathbf{25.61 \pm 0.593}$	27.82 ± 0.676
3	27.06 ± 0.259	28.34 ± 0.318	$\mathbf{23.08 \pm 0.174}$	25.66 ± 0.214
4	34.81 ± 0.173	34.90 ± 0.224	34.59 ± 0.189	$\mathbf{34.52 \pm 0.214}$
5	30.10 ± 0.720	28.30 ± 0.274	$\mathbf{23.67 \pm 0.224}$	25.89 ± 0.218
6	20.69 ± 0.634	23.14 ± 0.542	$\mathbf{16.48 \pm 0.241}$	18.75 ± 0.351
7	2.90 ± 0.112	3.79 ± 0.226	$\mathbf{2.24 \pm 0.123}$	3.03 ± 0.246
8	7.98 ± 0.660	2.66 ± 0.079	$\mathbf{2.49 \pm 0.074}$	3.02 ± 0.164
9	14.63 ± 0.324	7.43 ± 0.373	$\mathbf{4.84 \pm 0.156}$	6.71 ± 0.269
10	4.32 ± 0.235	6.48 ± 0.350	$\mathbf{4.00 \pm 0.194}$	6.11 ± 0.347
11	24.18 ± 0.193	26.53 ± 0.127	$\mathbf{22.08 \pm 0.085}$	22.22 ± 0.100
12	$\mathbf{3.09 \pm 0.127}$	5.21 ± 0.555	3.73 ± 0.179	5.70 ± 0.679
13	12.80 ± 0.325	11.34 ± 0.180	$\mathbf{9.93 \pm 0.043}$	10.95 ± 0.256

approach over single-objective approaches for tackling the model selection problem. In order to statistically assess the performance of these two approaches over the suite of 13 benchmark data sets, the Wilcoxon signed rank test with a 95% of confidence was used. According to this test, MOMTS-S2 is statistically better than PSMS.

5 Conclusions

In this paper, we presented our research proposal on the full model selection problem. We proposed to approach it as a multi-objective optimization one. We have a general way for estimating experimentally the model complexity, by using the VC-dimension. Our formulation showed the following advantages: (i) the experimental way for measuring the VC dimension allows us to consider different learning algorithms in a general framework, and also allows making the method extensible to the full model selection problem; (ii) our proposal showed a competitive performance over different benchmark data sets, which makes it applicable to problems from diverse domains; and (iii) the multiple non-dominated solutions obtained through the multi-objective formulation facilitates its extension to ensembles of models.

The VC dimension is experimentally estimated, which implies that a model must be trained and tested a number of times. This makes it computationally expensive. As part of our future work, we want to explore the surrogate-assisted evolutionary computation to reduce the computational cost. We also want to extend our current approach to the full model selection problem, i.e., considering feature selection and data pre-processing into the model selection process. Studying more effective ways for constructing an ensemble (possibly) by using a second level of optimization would also be another interesting direction for this research work.

References

1. Ayat, N., Cheriet, M., Suen, C.: Automatic model selection for the optimization of SVM kernels. Pattern Recogn 38(10), 1733 – 1745 (2005)
2. Aydin, I., Karakose, M., Akin, E.: A multi-objective artificial immune algorithm for parameter optimization in support vector machine. Appl Soft Comput 11(1), 120 – 129 (2011)
3. Bao, Y., Hu, Z., Xiong, T.: A PSO and pattern search based memetic algorithm for SVMs parameters optimization. Neurocomputing 117(0), 98 – 106 (2013)
4. Chang, C.C., Lin, C.J.: Libsvm: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2(3), 27:1–27:27 (2011)
5. Chapelle, O., Vapnik, V., Bousquet, O., Mukherjee, S.: Choosing multiple parameters for support vector machines. Mach Learn 46(1-3), 131–159 (2002)
6. Chatelain, C., Adam, S., Lecourtier, Y., Heutte, L., Paquet, T.: A multi-model selection framework for unknown and/or evolutive misclassification cost problems. Pattern Recogn 43(3), 815 – 823 (2010)
7. Coello Coello, C.A., Lamont, G.B., Veldhuizen, D.A.V.: Evolutionary Algorithms for Solving Multi-Objective Problems. Genetic and Evolutionary Computation, Springer US, 2 edn. (2007)
8. Corne, D.W., Jerram, N.R., Knowles, J.D., Oates, M.J., J, M.: Pesa-ii: Region-based selection in evolutionary multiobjective optimization. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO2001. pp. 283–290. Morgan Kaufmann Publishers (2001)
9. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE T Evol Comput 6(2), 182–197 (2002)
10. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res 7, 1–30 (2006)
11. Escalante, H.J., Montes, M., Sucar, L.E.: Particle swarm model selection. J Mach Learn Res 10, 405–440 (2009)
12. Gomes, T.A.F., Prudencio, R.B.C., Soares, C., Rossi, A.L.D., Carvalho, A.: Combining meta-learning and search techniques to svm parameter selection. In: Proceedings of the 11^{th} Brazilian Symposium on Neural Networks. pp. 79–84 (2010)
13. Gorissen, D., Dhaene, T., Turck, F.D.: Evolutionary model type selection for global surrogate modeling. J Mach Learn Res 10, 2039–2078 (2009)
14. Li, W., Liu, L., Gong, W.: Multi-objective uniform design as a svm model selection tool for face recognition. Expert Syst Appl 38(6), 6689 – 6695 (2011)
15. Miranda, P.B.C., Prudencio, R.B.C., Carvalho, A.C.P.L.F., Soares, C.: Multi-objective optimization and meta-learning for svm parameter selection. In: Neural Networks (IJCNN), The 2012 International Joint Conference on. pp. 1–8 (2012)
16. Rätsch, G., Onoda, T., Müller, K.R.: Soft margins for adaboost. Mach Learn 42(3), 287–320 (2001)
17. Reif, M., Shafait, F., Dengel, A.: Meta-learning for evolutionary parameter optimization of classifiers. Mach Learn 87(3), 357–380 (2012)
18. Rosales-Pérez, A., Escalante, H.J., Gonzalez, J.A., Reyes, C.A.: Bias and variance optimization for svms model selection. In: The 26^{th} FLAIRS Conference (2013)
19. Rosales-Pérez, A., Escalante, H.J., Gonzalez, J.A., Reyes-Garcia, C.A., Coello Coello, C.A.: Bias and variance multi-objective optimization for support vector machines model selection. In: Sanches, J.a.M., Micó, L., Cardoso, J.S. (eds.) Pattern Recognition and Image Analysis. LNCS, vol. 7887, pp. 108–116. Springer Berlin Heidelberg (2013)

20. Suttorp, T., Igel, C.: Multi-objective optimization of support vector machines. In: Jin, Y. (ed.) Multi-Objective Machine Learning, Studies in Computational Intelligence, vol. 16, pp. 199–220. Springer Berlin / Heidelberg (2006)
21. Valentini, G., Dietterich, T.G.: Bias-variance analysis of support vector machines for the development of SVM-based ensemble methods. J Mach Learn Res 5, 725–775 (2004)
22. Vapnik, V., Levin, E., Le Cun, Y.: Measuring the VC-dimension of a learning machine. Neural Comput 6(5), 851–876 (1994)
23. Wolpert, D.H.: The lack of a priori distinctions between learning algorithms. Neural Comput 8(7), 1341–1390 (1996)
24. Zhang, Q., Li, H.: MOEA/D: A multiobjective evolutionary algorithm based on decomposition. IEEE T Evol Comput 11(6), 712–731 (2007)