A Multi-objective Optimization Design Framework for Ensemble Generation

Victor Henrique Alves Ribeiro
Industrial and Systems Engineering Graduate Program
Pontifical Catholic University of Paraná (PUCPR)
Curitiba, Paraná, Brazil
victor.henrique@pucpr.edu.br

Gilberto Reynoso-Meza
Industrial and Systems Engineering Graduate Program
Pontifical Catholic University of Paraná (PUCPR)
Curitiba, Paraná, Brazil
g.reynosomeza@pucpr.br

ABSTRACT

Machine learning algorithms have found to be useful for the solution of complex engineering problems. However, due to problem's characteristics, such as class imbalance, classical methods may not be formidable. The authors believe that the application of multi-objective optimization design can improve the results of machine learning algorithms on such scenarios. Thus, this paper proposes a novel methodology for the creation of ensembles of classifiers. To do so, a multi-objective optimization design approach composed of two steps is used. The first step focus on generating a set of diverse classifiers, while the second step focus on the selection of such classifiers as ensemble members. The proposed method is tested on a real-world competition data set, using both decision trees and logistic regression classifiers. Results show that the ensembles created with such technique outperform the best ensemble members.

CCS CONCEPTS

• **Computing methodologies → Ensemble methods**; *Continuous space search*; Supervised learning by classification;

KEYWORDS

Ensemble methods, multi-objective optimization, logistic regression, decision trees.

ACM Reference format:

Victor Henrique Alves Ribeiro and Gilberto Reynoso-Meza. 2018. A Multi-objective Optimization Design Framework for Ensemble Generation. In *Proceedings of Genetic and Evolutionary Computation Conference Companion, Kyoto, Japan, July 15–19, 2018 (GECCO '18 Companion),* 4 pages.
https://doi.org/10.1145/3205651.3208219

1 INTRODUCTION

Machine learning algorithms have been found to be useful when dealing with complex engineering problems, such as the development of soft sensors and fault detection systems [7]. However, classical machine learning methods are usually optimized by means of its global accuracy, and there can be a lack of performance when applying such techniques to solve imbalanced classification problems, where we have a majority class, with a large number of examples, and a minority class, with few examples [5]. In such scenario, the global accuracy is not enough to indicate a good predictor, since even considering all examples as the majority class, a high accuracy is achieved.

Multi-objective machine learning [6] is a useful tool for solving such complex problems. By approximating the trade-off between the conflicting objectives of a problem, such as the majority and minority class-specific accuracies, a preferable solution can be selected based on the decision maker preferences. Diverse ensemble generation, or multi-objective ensemble generation [3], is indicated as one of the approaches for Pareto based supervised learning. Such technique focuses on generating ensembles composed of the diverse models achieved using a multi-objective optimization design (MOOD) approach.

Multi-objective ensemble generation has been the subject of study by many authors, and much work has been done on the definition of objectives for optimization, such as the trade-off between: accuracy and complexity [3, 6, 15]; mean squared error and the number of connections of recurrent neural networks [16]; accuracy and different diversity measures [4]. However, it is indicated in [3] that the significance of the multi-objective approach for ensemble generation will be considerably reduced without a clear idea of how to select a subset of the Pareto optimal solutions, and much work remains to be done along this line of research. The selection of members close to the Pareto front's "knee point" has been performed in [3, 16], while different methods for member selection have been tested in [14, 15].

Thus, this paper proposes a novel MOOD methodology for the creation of such ensembles, where the problem is split into the following two multi-objective problems (MOP): the first step focus on creating a set of diverse non-dominated classifiers, by selecting different features and model parameters; the second step is focused on selecting the previous classifiers as ensemble members. Both problems are optimized with a multi-objective optimization (MOO) algorithm, and a multi-criteria decision making (MCDM) step is applied in order to select a preferable final ensemble after the last MOP.

The remainder of this paper is presented as follows: section 2 presents the background necessary for multi-objective ensemble generation; section 3 presents the proposed methodology; section 4 presents the experiment and its results; and section 5 presents the conclusion and final remarks on the subject.

GECCO '18 Companion, July 15–19, 2018, Kyoto, Japan

ACM ISBN 978-1-4503-5764-7/18/07…$15.00
https://doi.org/10.1145/3205651.3208219

2 BACKGROUND

In order to create optimized classifiers for complex engineering classification problems, a novel methodology for the generation of ensembles of classifiers, using a MOOD approach, is proposed. Thus, this topic presents a background on classification task, ensemble methods and MOOD.

2.1 Classification

The classification task is an instance of supervised learning, where a classifier must predict the category of a new observation, given a training data of multiple observations with known categories. It is typically composed of three components: a training set, an inducer and a classifier, where the inducer is a learning algorithm that creates a classifier for a specific training set [13]. Two learning algorithms are used in this work, the logistic regression (LR) [8] and the decision trees (DT) [11].

LR is a two-class linear separator, where the sigmoid function is used to compute the probability of an observation belonging to a certain category, given the model's trained weights and the observation's features. Equation 1 shows such procedure, where C_1 represents the true class, σ the sigmoid function, $\boldsymbol{\theta}$ the vector of trained weights and $\boldsymbol{X}$ the observation's vector of features.

$$p(C_1|\boldsymbol{X}) = \sigma(\boldsymbol{\theta}^T\boldsymbol{X}) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T\boldsymbol{X}}} \tag{1}$$

DT, on the other hand, are models that perform sequential decision making, where a tree structure is created to subdivide the variables space into separate regions, each one related to one category. Figure 1 illustrates such structure.

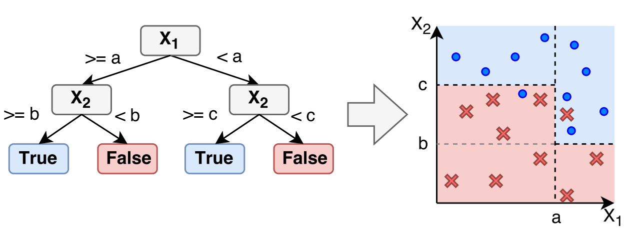

Figure 1: Decision Tree (DT) structure (left) creating rules to divide the features space into separate categories (right) for classification.

2.2 Ensemble Methods

The ensemble methodology consists on the combination of different classifiers, resulting in a model that outperforms its members. A rich review on the subject is presented in [13] for interested readers. Additionally, a review on multi-objective ensemble generation is presented in [3].

In order to build ensembles of classifiers, the following two steps are necessary:

- **Member generation:** diverse classifiers are trained by manipulating the training data set and/or changing the structure of the inducers, by adjusting hyperparameters or using different learning algorithms;
- **Member combination:** a technique, such as majority voting or average weight, is used to combine the members' prediction outputs.

Figure 2 presents an ensemble, which is composed of diversely generated classifiers and a combiner component.

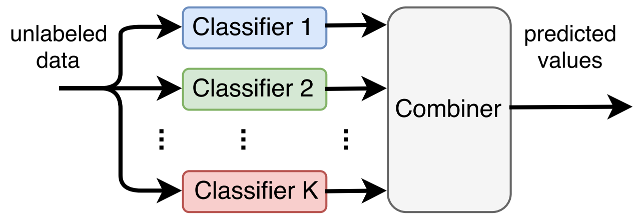

Figure 2: Structure of an ensemble of classifiers.

Ensemble methods are known for improving the accuracy of classifiers [13]. It has also been widely used in the literature [3, 4, 6, 14–16] to solve classification problems. Thus, such technique is applied in the proposed work.

2.3 Multi-Objective Optimization Design

According to [10], the MOOD approach can be divided into three steps:

- **MOP definition:** the problem's objectives, decision variables and constraints are defined;
- **MOO:** a search algorithm is used to retrieve a Pareto front approximation, a set composed of many non-dominated solutions with conflicting objectives. Figure 3 illustrates the Pareto front for a generic MOP.
- **MCDM:** visualization and ranking techniques are applied to aid the decision maker in the task of finding a preferable final solution among the Pareto front.

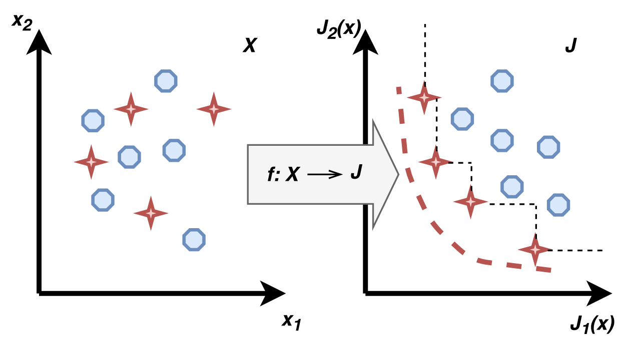

Figure 3: Representation of the decision variables (left) and the objectives (right) for a multi-objective problem (MOP), where the blue circles represent the dominated solutions while the red stars represent the non-dominated solutions that approximate a Pareto front (red traced curve).

3 METHODOLOGY

In order to effectively build an ensemble of classifiers, the present work separates the ensemble generation into two different steps:

- **Member generation:** the generation of a set of diverse Pareto optimal classifiers as candidate ensemble members;
- **Member selection:** the selection of the classifiers that will be part of the final ensemble.

For the member generation step, the following MOP is stated for both classifiers[1]:

$$\min_{\boldsymbol{x}} \boldsymbol{J_g}(\boldsymbol{x}) = [-J_1(\boldsymbol{x}), -J_2(\boldsymbol{x}), -J_3(\boldsymbol{x}), -J_4(\boldsymbol{x}), J_5(\boldsymbol{x})] \tag{2}$$

subject to:

$$x_i \in \{0, 1\}, \qquad i = [1, \ldots, n] \tag{3}$$

$$-20 < x_j < 20, \quad j = [n+1, \ldots, 2n] \tag{4}$$

where the objectives are: global accuracy ($J_1(\boldsymbol{x})$ [dimensionless]); true positive rate ($J_2(\boldsymbol{x})$ [dimensionless]); true negative rate ($J_3(\boldsymbol{x})$ [dimensionless]); F1 score, or the harmonic mean of sensitivity and precision ($J_4(\boldsymbol{x})$ [dimensionless]); and classifier's complexity ($J_5(\boldsymbol{x})$), defined as absolute sum of weights for logistic regression or number of features for decision tree. The decision variables are: selection of each of the n features (x_i), both for LR and DT; and the initial weight assigned to each of the n features (x_j), used only for LR.

Next, the member selection step MOP is defined for both learning algorithms as follows:

$$\min_{\boldsymbol{x}} \boldsymbol{J_s}(\boldsymbol{x}) = [-J_1(\boldsymbol{x}), -J_2(\boldsymbol{x}), -J_3(\boldsymbol{x}), -J_4(\boldsymbol{x}), J_6(\boldsymbol{x})] \tag{5}$$

subject to:

$$x_k \in \{0, 1\}, k = [1, \ldots, m] \tag{6}$$

where the new objective is the ensemble's complexity ($J_6(\boldsymbol{x})$), defined as the sum of the members' complexities. The decision variables (x_k) define the ensemble membership for each of the m resulting classifiers from the previous step.

The spherical pruning multi-objective differential evolution (sp-MODE) [12] is used for both the member generation and member selection steps. The algorithm is configured with a scaling factor of 0.5, a crossover probability of 0.5, a population of 50 individuals, and a maximum number of 200 generations or 10000 function evaluations.

Finally, in order to select a preferable ensemble, physical programming [9] is used in order to rank the ensembles based on the preferences listed in Table 1.

As a conclusion, the methodology, illustrated in Figure 4, consists in: a member generation step, where training data and an inducer are used to create a set of Pareto optimal classifiers; and the member selection step, where the resulting classifiers are used to create a set of non-dominated ensembles, and a MCDM step selects the preferable final solution.

[1] The constraints from Equation 4 are not used for the creation of decision trees.

Table 1: Preference matrix for the member selection problem. Five scaled preference ranges have been defined: highly desirable (HD), desirable (D), tolerable (T) undesirable (U) and highly undesirable (HU).

Preference Matrix						
	← HD →	← D →	← T →	← U →	← HU →	
Objective	S_i^0	S_i^1	S_i^2	S_i^3	S_i^4	S_i^5
$J_1(\boldsymbol{x})$	0.00	0.10	0.20	0.40	0.60	1.00
$J_4(\boldsymbol{x})$	0.00	0.10	0.20	0.40	0.60	1.00
$J_6(\boldsymbol{x})$	0.00	0.10	0.20	0.40	0.60	1.00

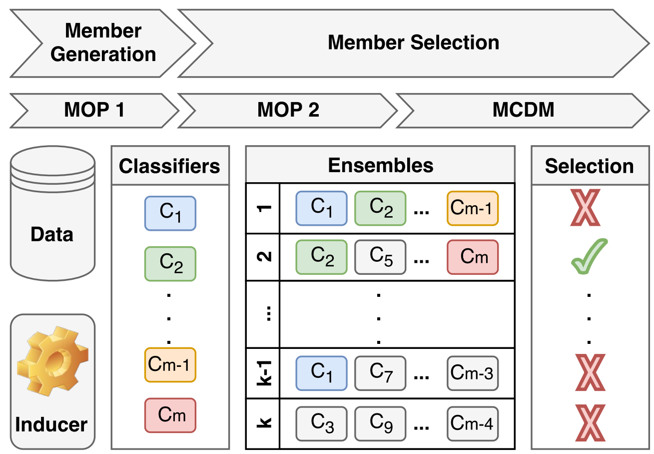

Figure 4: The proposed multi-objective optimization design (MOOD) framework for ensemble generation.

4 EXPERIMENT AND RESULTS

The data sets provided in SPOTSeven's GECCO challenge 2017 [1] have been used for training and testing the ensembles created with the proposed methodology. The training data set is split into training (70%) and validation (30%) sets for hold-out validation during the multi-objective optimization stage, while the test data set is used for analyzing final results. In total, 81 runs were executed to analyze the creation of ensembles of LR and DT classifiers.

The distribution plot of the F1 score for each classifier is shown in Figure 5. The first column presents the results for the best LR model for each run, where we can see a distribution with median value close to 0.3, maximum value of more than 0.4 and minimum value of less than 0.1. The second column presents the results for the selected ensembles of LR classifiers (MOEG-LR) for each run, where we can see a distribution with median value close to 0.45, maximum value of more than 0.5 and minimum value close to 0.2. The third column presents the results for the best DT for each run, where we can see a distribution with median value close to 0.45, maximum value of more than 0.55 and minimum value close to 0.45. The last column presents the results for the selected ensemble of DT (MOEG-DT) for each run, where we can see a distribution with median value close to 0.5, maximum value close to 0.55 and minimum value close to 0.45.

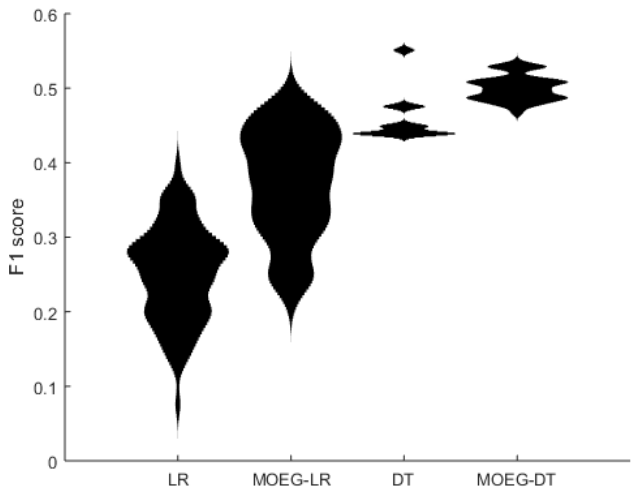

Figure 5: Distribution plot of the four classifiers.

The Friedman's test [2] is performed with such data, resulting on the mean ranks presented in Table 2. A small p–value of $3.2e^{-44}$ indicates that each model presents different results, while the classifier's mean ranks indicate that MOEG-DT achieved the best result, followed by DT, MOEG-LR and LR.

Table 2: Mean ranks result from the Friedman's test.

Classifier	Mean rank
Logistic Regression	3.9012
Multi-objective Ensemble Generation with Logistic Regression	2.9383
Decision Tree	2.0123
Multi-objective Ensemble Generation with Decision Tree	1.1481

The results can also be compared to the competition winner's results [1], where a F1 score of 0.44 has been achieved. The proposed framework is capable of outperform such results, having the MOEG-LR and the MOEG-DT classifiers a median F1 score of 0.45 and 0.5, respectively.

Such results indicate that the proposed MOOD framework for ensemble generation grants a performance improvement for both the LR and DT classifiers. Also, it presents a competitive algorithm to be used in real-world applications and benchmark problems.

5 CONCLUSIONS

In order to improve classification performance of learning algorithms, this paper presents a methodology for the automatic creation of classification ensembles. The proposed method is composed of two MOO stages: creation of a Pareto optimal set of ensemble candidate members; and, selection of such candidates to create an optimal final ensemble. As a solution to the problem indicated by [3], such methodology can improve the significance of the multi-objective approach for the creation of ensembles.

Results show that the proposed methodology improves the performance of LR and DT classifiers. Also, such methodology has been applied to a challenge data set, achieving results that outperforms the challenge's winner. Thus, multi-objective ensemble generation is indicated to solve complex real-world classification problems.

Future work can focus on: the application of different learning algorithms; different MOO algorithms; and different MCDM techniques to find a preferable ensemble. It is also recommended to test different ensemble combination techniques. Finally, the authors intend to compare such methodology with other ensemble generation techniques in the future, using benchmark data sets.

ACKNOWLEDGMENTS

This work is supported by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), through the grant PROSUC/159063/2017-0, and is under the research initiative *Multi-objective optimisation design (MOOD) procedures for engineering systems: Industrial applications, unmanned aerial systems and mechatronic devices*, supported by the National Council of Scientific and Technological Development of Brazil (CNPq) through the grant PQ-2/304066/2016-8.

REFERENCES

[1] Sowmya Chandrasekaran, Martina Friese, JÃűrg Stork, Margarita Rebolledo, and Thomas Bartz-Beielstein. 2017. GECCO Challenge 2017 Monitoring of drinking-water quality. (2017). http://www.spotseven.de/gecco/gecco-challenge/gecco-challenge-2017/
[2] Wayne W Daniel et al. 1978. *Applied nonparametric statistics.* Houghton Mifflin.
[3] Shenkai Gu, Ran Cheng, and Yaochu Jin. 2015. Multi-objective ensemble generation. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 5, 5 (2015), 234–245.
[4] Shenkai Gu and Yaochu Jin. 2014. Generating diverse and accurate classifier ensembles using multi-objective optimization. In *Computational Intelligence in Multi-Criteria Decision-Making (MCDM), 2014 IEEE Symposium on.* IEEE, 9–15.
[5] Nathalie Japkowicz and Shaju Stephen. 2002. The class imbalance problem: A systematic study. *Intelligent Data Analysis* 6, 5 (2002), 429–449. https://doi.org/10.1.1.711.8214
[6] Yaochu Jin and Bernhard Sendhoff. 2008. Pareto-based multiobjective machine learning: An overview and case studies. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 38, 3 (2008), 397–415.
[7] Petr Kadlec, Bogdan Gabrys, and Sibylle Strandt. 2009. Data-driven Soft Sensors in the process industry. *Computers and Chemical Engineering* 33, 4 (2009), 795–814. https://doi.org/10.1016/j.compchemeng.2008.12.012
[8] Scott Menard. 2018. *Applied logistic regression analysis.* Vol. 106. SAGE publications.
[9] Achille Messac. 1996. Physical programming: effective optimization for computational design. *AIAA journal* 34, 1 (1996), 149–158.
[10] Gilberto Reynoso Meza, Xavier Blasco Ferragud, Javier Sanchis Saez, and Juan Manuel Herrero Durá. 2016. *Controller tuning with evolutionary multiobjective optimization: a holistic multiobjective optimization design procedure.* Vol. 85. Springer.
[11] Brijain R Patel and Kushik K Rana. 2014. A Survey on Decision Tree Algorithm For Classification. *Ijedr* 2, 1 (2014), 1–5.
[12] Gilberto Reynoso-Meza, Javier Sanchis, Xavier Blasco, and Sergio García-Nieto. 2014. Multiobjective evolutionary algorithms for multivariable PI controller tuning. *Applied Soft Computing* 24 (2014), 341 – 362.
[13] Lior Rokach. 2010. Ensemble-based classifiers. *Artificial Intelligence Review* 33, 1 (2010), 1–39.
[14] Alejandro Rosales-Perez, Salvador Garcia, Jesus A Gonzalez, Carlos A Coello Coello, and Francisco Herrera. 2017. An Evolutionary Multi-Objective Model and Instance Selection for Support Vector Machines with Pareto-based Ensembles. *IEEE Transactions on Evolutionary Computation* PP, 99 (2017), 1–1.
[15] Alejandro Rosales-Pérez, Jesus A Gonzalez, Carlos A Coello Coello, Hugo Jair Escalante, and Carlos A Reyes-Garcia. 2014. Multi-objective model type selection. *Neurocomputing* 146 (2014), 83–94.
[16] Christopher Smith and Yaochu Jin. 2014. Evolutionary multi-objective generation of recurrent neural network ensembles for time series prediction. *Neurocomputing* 143 (2014), 302–311.