

PAPER • OPEN ACCESS

An Effective Metaheuristic for Bi-objective Feature Selection in Two-Class Classification Problem

To cite this article: A A Lyubchenko *et al* 2019 *J. Phys.: Conf. Ser.* **1210** 012086

View the [article online](#) for updates and enhancements.

You may also like

- [A hybrid genetic algorithm for solving bi-objective traveling salesman problems](#)
Mei Ma and Hecheng Li
- [Bi-criteria Pareto optimization to balance irradiation time and dosimetric objectives in proton arc therapy](#)
Sophie Wuyckens, Lewei Zhao, Michael Saint-Guillain *et al.*
- [A review and experimental study on the application of classifiers and evolutionary algorithms in EEG-based brain-machine interface systems](#)
Farajollah Tahermezhad-Javazm, Vahid Azimirad and Maryam Shoaran



UNITED THROUGH SCIENCE & TECHNOLOGY

 **The Electrochemical Society**
Advancing solid state & electrochemical science & technology

**248th
ECS Meeting**
Chicago, IL
October 12-16, 2025
Hilton Chicago

**Science +
Technology +
YOU!**

**SUBMIT
ABSTRACTS by
March 28, 2025**

SUBMIT NOW

The advertisement features a central image of a smiling woman with long dark hair, wearing a brown blazer, gesturing with her hands. The background is a dark blue with a network of white lines and nodes, suggesting a scientific or technological theme. The banner is framed by decorative borders at the top and bottom consisting of repeating circular patterns with diagonal lines.

An Effective Metaheuristic for Bi-objective Feature Selection in Two-Class Classification Problem

A A Lyubchenko¹, J A Pacheco¹, S Casado¹, L Nuñez²

¹University of Burgos, Burgos, Spain

²IE Business School, Madrid, Spain

e-mail: allyubchenko@gmail.com

Abstract. Feature selection is known as a very useful technique in machine learning practice as it may result in the development of more straightforward models with better accuracy. Traditionally, feature selection is considered as a single-objective problem, however, it can be easily formulated in terms of two objectives. The solving of such problems requires the application of appropriate multi-objective optimization methods that do not always offer equally good solutions even under the same conditions. This paper focuses on the development of a metaheuristic optimization approach for bi-objective feature selection problem in two-class classification. We consider the solving of this problem in terms of minimization of both misclassification error and feature subset size. For solving the considered problem, an adaptation of the Multi-Objective Adaptive Memory Programming (MOAMP) metaheuristic based on the tabu search strategy is proposed. Our MOAMP adaption has been utilized to obtain the sets of most relevant features for two real classification problems with two classes. Finally, using popular Pareto front quality indicators, the obtained results have been compared with the sets of non-dominated solutions derived by the well-known NSGA2 algorithm. The conducted research allows concluding about the ability of the MOAMP adaptation to get a better efficient frontier for the same number of objective function calls.

Keywords: bi-objective feature selection, classification, tabu search, MOAMP

1. Introduction

Feature or variable selection (FS) constitutes a process of selecting the most relevant variables that form a less redundant subset of features with keeping their explanatory power with respect to the target variable. FS is a very popular methodology in machine learning as it allows increasing the prediction accuracy, constructing less complex predictor with faster training process, getting a better understanding of the analyzed data [1], as well as it may contribute to the prevention of model overfitting [2]. However, feature selection is known as a problem with high computational costs that results from the exponential growth of the number of possible feature subsets with an increase of a total number of variables [3]. Hence, taking into account the modern tendency of data dimensionality growth, the development of effective FS techniques remains to be a relevant problem which solving can contribute significantly in machine learning practice.

For minimization of the size of feature subset and keeping the maximum of the explanatory ability of a predictor, FS should be considered as a bi-objective optimization problem. In particular, FS problem statement, in terms of feature subset size minimization and classification accuracy maximization, is seen popular [4, 5].

Classification is a type of supervised machine learning problem that lies in predicting to which category a new instance belongs to using a previously trained classifier on the basis of available observations. The development of robust classifiers with optimal feature subsets is a widespread



challenging problem, which covers a variety of real-life applications such as text data processing [6], facial recognition technology [7], faults detection in the technical system [8], bankruptcy prediction [9] etc.

This paper is focused on the development of an effective feature selection approach for pervasive two-class classification problem. We deal with the selection of relevant feature subset from the bi-objective perspective, considering both minimization of misclassification error and features subset size. Consequently, the objective is to obtain a set of non-dominated solutions forming an approximation of Pareto front for a considered classification problem. For doing so, we have proposed a methodology based on an adaptation of Multi-objective Adaptive Memory Programming (MOAMP) metaheuristic [10, 11] that uses the tabu search strategy for getting global optima [12, 13]. The MOAMP approach has been used successfully in some real problems of scheduling-routing planning [14-16] and supply chain design [17]. In this work, an attempt has been made to develop an adaptation of MOAMP metaheuristic with respect to the bi-objective feature selection problem.

For the purpose of reasonable merit assessment, results of the MOAMP adaptation are compared with the approximations obtained by widely recognized Non-Dominated Sorting Genetic Algorithm 2 (NSGA2). The comparison is based on the estimation of Pareto front quality indicators such as cardinality and hypervolume. Both multi-objective optimization techniques are applied to features selection problem for two popular classifiers based on Logistic regression and Linear Discriminant analysis. The experiments have been performed using on real-life financial data. But the proposed MOAMP based bi-objective feature selection approach can be easily applied to different datasets that will be an object of our future research.

The remainder of the paper is organized as follows. Section 2 presents the bi-objective feature selection problem. In section 3, the proposed MOAMP based technique is described. Section 4 is devoted to the experimental part and comparative analysis of MOAMP and NSGA2 results. Finally, in section 5, the main conclusions are offered.

2. Problem formulation

In case of a classification problem, the objective of machine learning is the training of a classifier using an available set A of n observations (or cases). Each observation is represented by a vector of values of m features $V = \{v_1, v_2, \dots, v_m\}$ and an information about to which class (0 or 1) the case belongs to (also named "label information"). Hence, in supervised learning, the training set A consists of a matrix with n rows and m columns, which corresponds to each feature value for each observation, and a vector L of label information for all n available observations. Generally, the training set A is divided into two parts, the first is in fact for the training process and the second subset serves for model validation. The proven practice to overcome the classifier overfitting is to perform cross-validation. After the training process with cross-validation, the classifier can be used for prediction of the label for a new vector of feature values V .

Thus, the aim of feature selection is to find the smallest possible feature subset S of size k , $S \subset V$, that keeps the predictable ability of the classifier. In other words, a classifier built on the subset S , must predict the labels of the available observations with minimal misclassification error. Consequently, the bi-objective feature selection problem can be formulated as the search such a feature subset S with the minimum size (objective function $f_1(S)$) that concurrently minimize the misclassification error of the considered classifier (objective function $f_2(S)$). As we deal with two objectives, our challenge is to find a set of non-dominated solution S_{ND} that forms efficient frontier or Pareto front of the considered bi-objective optimization problem.

3. Bi-objective feature selection metaheuristic

The Multi-objective Adaptive Memory Programming (MOAMP) metaheuristic based on the tabu search (TS) strategy tries to adapt TS routine to the structure of the set of non-dominated solutions of a multi-objective problem [18]. Generally, the MOAMP framework includes three phases. The first phase is intended to get an initial approximation of the efficient set S_{ND} performing series of tabu searches. Whereas the second phase tries to improve the previously obtained set of solutions

conducting searches with a modified objective function. And on the last stage, the searches are carried out around all solutions from the current efficient set S_{ND} until these searches does not provide improvements of the current efficient set. It is should be noticed that the MOAMP metaheuristic is quite a flexible technique and it can be considered as a framework that should be adapted to each particular type of an optimization problem. Here, in pseudo-code 1, we present our adaptation of MOAMP concept for bi-objective feature selection problem with some comments for each step.

Pseudo-code 1. Adaptation of MOAMP metaheuristic

Phase 1

1. Set $S_{ND} = \emptyset$
2. Tabu searches $f_1^{min} \rightarrow f_2^{min}$
 - 2.1 $S_1 = \{v_r\}$ where $v_r = \operatorname{argmin}\{f_2(v_j): v_j \in V\}$;
 - 2.2 $S \leftarrow \operatorname{TabuSearch}(f_2, S_1)$;
3. Tabu searches $f_2^{min} \rightarrow f_1^{min}$
 - 3.1 $S_2 \leftarrow \operatorname{TabuSearch}(f_2, V)$;
 - 3.2 $S \leftarrow \operatorname{TabuSearch}(f_1, S_2)$
4. Update S_{ND} while performing this phase

Phase 2

Repeat

5. $\lambda = \operatorname{uniform}(0, 1)$;
6. $S \leftarrow \operatorname{TabuSearch}(F_\lambda, S)$;
7. Update S_{ND}

Until S_{ND} does not change for $\max\text{Phase2}$ iterations

Phase 3

Repeat

8. Explore the neighborhood of each S from S_{ND} that has not been explored and update S_{ND}
9. Identify new non-dominated solutions

Until S_{ND} does not change

The search process starts from the approximation the initial efficient set S_{ND} with the empty set (step 1). The core of the first phase is the execution of two consecutive tabu search procedures (step 2 and 3). Step 2 serves to find the left side optimum S_1 that corresponds to the minimum of feature subset size f_1^{min} and to the maximum of misclassification error f_2^{max} . Then, starting from the solution S_1 , *TabuSearch* routine is conducted, moving to the right side of solution space and trying to reach the minimum of misclassification error f_2^{min} .

Step 3 performs a similar search, but from the right side of the solution space. First, using the same *TabuSearch* routine with the complete set of variables V , it seeks to reach the right side optimum S_2 that coincides with the minimum of the misclassification error f_2^{min} and the corresponded maximum number of features f_1^{max} . Furthermore, the optimum S_2 is used like a starting point for *TabuSearch* procedure, which aims to move back to the left side optimum S_1 . Thus, in steps 2 and 3, two cross-searches are carried out on the basis of tabu strategy $f_1^{min} \rightarrow f_2^{min}$ and $f_2^{min} \rightarrow f_1^{min}$, consequently. While both steps are conducted, the current set of non-dominated solutions S_{ND} is continually updated with the new solutions that could improve it.

As already mentioned, the second phase of MOAMP metaheuristic is designed to improve the set of non-dominated solutions S_{ND} , obtained in Phase 1. This is possible through the change of the objective function that is used during tabu searches. The modified objective function is given by:

$$F_\lambda(S) = \max \left\{ \lambda \frac{f_1(S) - f_1^{min}}{f_1^{max} - f_1^{min}}, (1 - \lambda) \frac{f_2(S) - f_2^{min}}{f_2^{max} - f_2^{min}} \right\}, \quad (1)$$

where λ is the weight factor randomly generated from a uniform distribution $U(0, 1)$.

Consequently, in step 6, *TabuSearch* routine seeks to find the minimum of the objective function $F_\lambda(S)$ using the solution, reached in step 3.2 as a starting point, while each visited solution is considered to update S_{ND} . Phase 2 performs so many experiments until new searches do not offer improvements for the efficient set S_{ND} during the specified number of iterations $maxPhase2$.

The last third phase of the MOMAP adaptation does not perform optimization procedures, instead, the neighborhood of each solution from the efficient set S_{ND} is analysed. In other words, only neighbour solutions are added to the set S_{ND} if they improve it. This stage is continued until new explorations of neighbourhoods do not give better results.

The neighborhood exploration of the last MOAMP phase is similar to the construction of a candidate list in tabu searches. In both cases, three types of operations (or moves) are done: add, drop and interchange. It means that, in the case of the add operation, a variable from the set of all unused features can be appended to the current vector of variables. Drop move serves for the opposite operation that excludes a variable from the current vector and put it into the set of unused features. Interchange implies the combining of add and drop operations when one feature is removed from the current vector of variables, whereas an alternative variable is appended to the vector.

General condition for making these moves is inactive tabu status of a considered feature. The control of tabu status is realized through a memory structure that contains two registers for add and drop moves. For example, if an untapped variable is considered to be added to the current vector of features, then it must have an inactive tabu status in the corresponded register. Consequently, the algorithm checks register that is responsible for operation to be done. In the case of interchange moves, both registers are checked. If the operation is permitted i.e. tabu status of the variable is inactive, then the algorithm performs this move and updates the corresponded register, activating tabu status for some number of search iterations. This parameter is called tabu tenure and defines how long tabu status can be active.

We have mentioned the general condition to perform a move, and it comes down to the inactive status of a feature. However, there is an exception that is called Aspiration criteria and it implies an opportunity to make a move even with active tabu status. This is possible only if this move can improve the best ever found solution. Hence, due to the application of tabu memory structure and Aspiration criteria, tabu search algorithms are able to explore solution space beyond local optimality.

Thus, implementing all three phases of the tabu search based MOAMP adaptation, it is possible to obtain set of non-dominated solutions or efficient set that forms Pareto front for the considered bi-objective feature selection problem.

In order to provide a better understanding of how our MOAMP metaheuristic works, in figure 1 and 2, the results of three MOAMP stages are illustrated. These data are a part of the experimental results that will be presented in section 4.

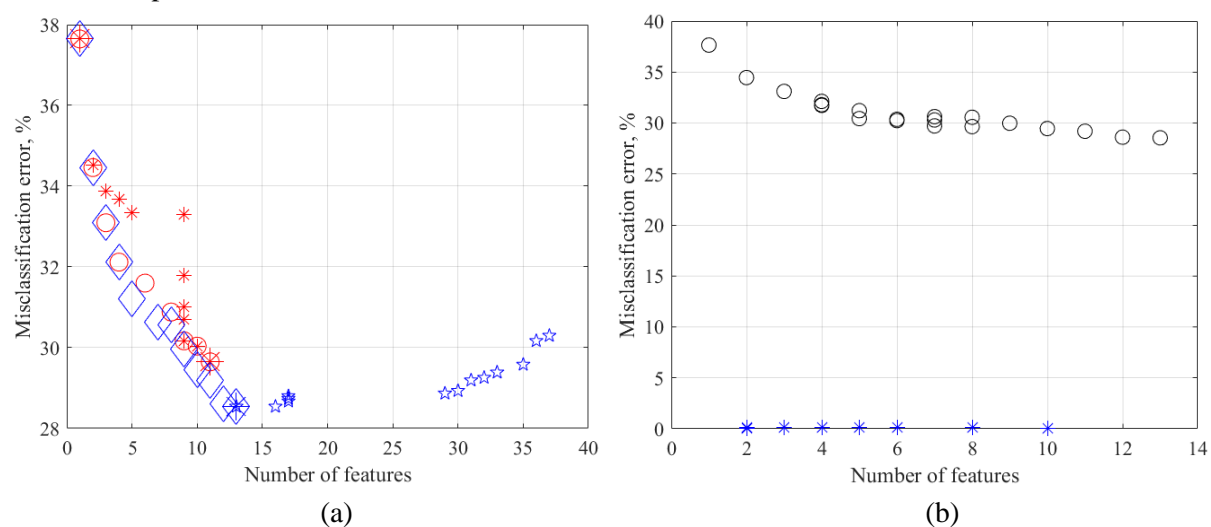


Figure 1. The results of the search process of Phase 1 (a) and Phase 2 (b).

Figure 1, a demonstrates some trajectories of tabu searches that were done during the first phase. The star-shaped marks correspond to the solutions, passed by *TabuSearch* routine in step 2.2, whereas the circle marks show solutions that were added to the current approximation of the efficient set S_{ND} on this step. The pentagram icons illustrate how the tabu search procedure was passing in step 3.1 in order to reach the optimum S_2 . The output of the first phase is the set of non-dominated solutions S_{ND} depicted by the diamond marks. This set was obtained in step 3.2 through updating current approximation with all visited solutions which could improve it.

In figure 1,b, we can see the improvement of the efficient set during a series of iterations thanks to the changed objective function $F_\lambda(S)$. The star-shaped marks correspond to the optima, reached by *TabuSearch* routine on different iterations.

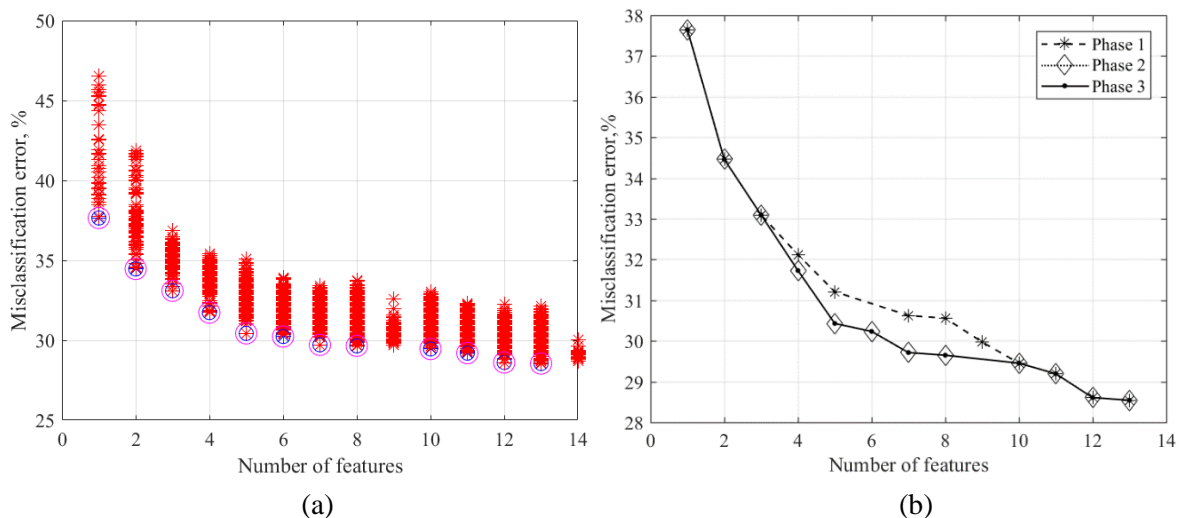


Figure 2. The results of the search process of Phase 3 (a) and all MOAMP stages (b).

The results of the last third phase are presented in figure 2, a. The star-shaped marks coincide with the process of neighborhood exploration for each solution in the current S_{ND} , whereas the circles define the solutions of the final set of non-dominated solutions S_{ND} . The solutions, derived through three MOAMP phase, are illustrated together in figure 2, b.

4. Experiments

In order to make the merit estimation of the proposed MOAMP adaptation, eight computational experiments have been performed with two real classification problems from the financial area, denoted S and I problem. Both datasets contain $m = 38$ features and $n = 1538$ observations. For each problem, two types of classifiers were considered, on the basis of Logistic regression (LR) and using Linear Discriminant analysis (LD). Consequently, four classification problems, S_{38_LR} , S_{38_LD} , I_{38_LR} , and I_{38_LD} , were investigated from the perspective of the abovementioned bi-objective features selection problem. For each stated classification problem, two experiments have been carried out, using the proposed MOAMP metaheuristic and well-known and proven NSGA2 algorithm. Both metaheuristics were driven for the same number of objective function calls and with the default seed of random number generator. All classifiers were trained with 5-folds cross-validation.

The MOAMP adaptation has three main hyperparameters, which values were set as follows: tabu tenure equals to $m-1$, $maxPhase2 = 5$ and the number of iterations of *TabuSearch* routine without improvement of the best solution equals to m . NSGA2 hyperparameters are given by: population size is $2 \cdot m$, crossover percentage is 0.7 and mutation rate that equals to $1/m$.

For each classification problem and metaheuristic, two Pareto front quality indicators were calculated on the basis of the obtained sets of non-dominated solutions, such as cardinality [19] and hypervolume [20]. Hence, these indicators are used for the comparison MOAMP and NSGA2 algorithms, and the greater value of an indicator, the more effective is the considered technique.

The graphical illustration of the obtained efficient sets for four classification problems is demonstrated in figure 3. For visual comparison, each graph presents the results of the last third phase of the MOAMP adaptation and the solutions sets of NSGA2 metaheuristic.

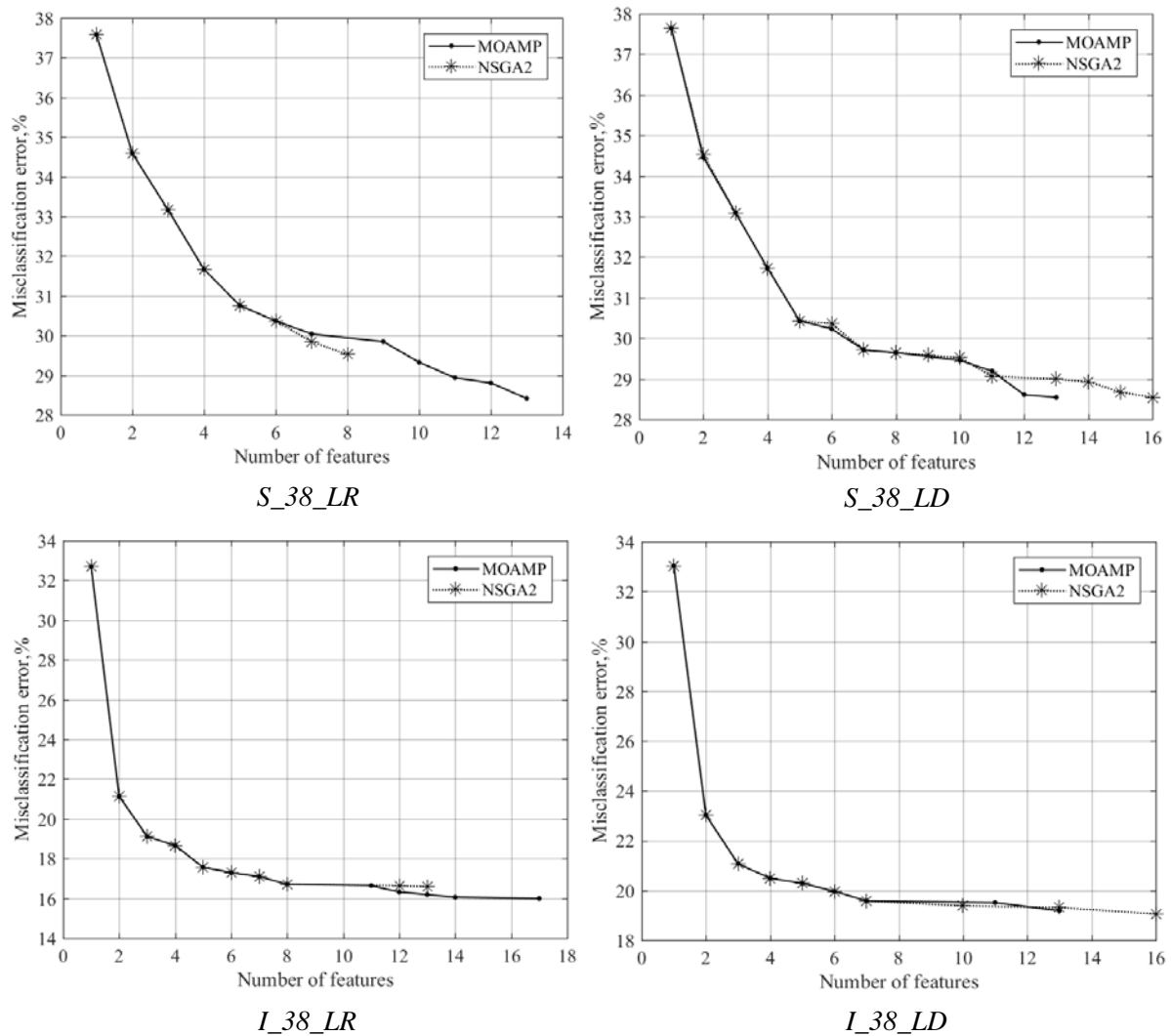


Figure 3. The experimental results for four considered classification problems.

In table 1, the results of cardinality and hypervolume indicators calculation are grouped. The estimations are presented for each classification problem and separately for MOAMP and NSGA2.

Table 1. Comparison of Pareto front indicators

Classification problem	Cardinality		Hypervolume	
	MOAMP	NSGA2	MOAMP	NSGA2
S_38_LR	0.8333	0.6667	0.7337	0.7223
S_38_LD	0.8462	0.6154	0.8044	0.7940
I_38_LR	1.0000	0.6154	0.9006	0.8911
I_38_LD	0.8000	0.9000	0.9065	0.9088

As we can see in table 1, in most cases the proposed MOAMP adaptation allows getting a better efficient set among the considered classification problems. Only in the case of *I_38_LD* problem, the NSGA2 approach shows moderately better results.

5. Conclusions

Feature selection remains to be a sought-after technique in machine learning as it helps to overcome some difficulties that can arise during the development of a robust prediction model. The consideration of variable selection problem from bi-objective perspective is natural and understandable. However, it requires the application of multi-objective optimization techniques that do not always provide equally good solutions even under the same conditions. In this research, we propose an adaptation the Multi-Objective Adaptive Memory Programming (MOAMP) metaheuristic for solving bi-objective feature selection problem. This adaptation is based on tabu search that allows reaching of a global optimum.

Using the developed adaptation, a series of computational experiments has been conducted for real two-class classification problems. MOAMP metaheuristic helped to solve bi-objective feature selection for two classifiers, based on logistic regression and linear discriminant analysis. For the purpose of merit estimation of the proposed approach, similar experiments have been carried out using the well-known and proven NSGA2 algorithm. The results of both metaheuristics were compared according to the cardinality and hypervolume indicators. The comparative analysis has shown the ability of the MOAMP adaptation to get a better efficient frontier for the same number of objective function calls.

Acknowledgments

This work was partially supported by FEDER funds and the Spanish Ministry of Economy and Competitiveness (Projects ECO2013-47129-C4-3-R and ECO2016-76567-C4-2-R), the Regional Government of “Castilla y León”, Spain (Project BU329U14) and the Regional Government of “Castilla y León” and FEDER funds (Project BU062U16); all of whom are gratefully acknowledged.

References

- [1] Reunanen J 2003 Overfitting in making comparisons between variable selection methods *Journal of Machine Learning Research* vol 3 pp 1371-1382
- [2] Mousin L et al. 2016 Feature Selection Using Tabu Search with Learning Memory: Learning Tabu Search *Learning and Intelligent Optimization Conference LION 10* <https://hal.archives-ouvertes.fr/hal-01370396>
- [3] Cotta C Sloper C Moscato P 2004 Evolutionary Search of Thresholds for Robust Feature Set Selection: Application to the Analysis of Microarray Data *EvoWorkshops. Applications of Evolutionary Computing* vol 3005 pp 21-30
- [4] Hancer E et al. 2018 Pareto front feature selection based on artificial bee colony optimization *Information Sciences* vol 422 pp 462-479 DOI: 10.1016/j.ins.2017.09.028
- [5] Hancer E Xue B Zhang M 2018 Differential evolution for filter feature selection based on information theory and feature ranking *Knowledge-Based Systems* vol 140 pp 103-119 DOI: 10.1016/j.knsys.2017.10.028
- [6] Rehman A et al. 2015 Relative discrimination criterion – A novel feature ranking method for text data *Expert Systems with Applications* vol 42(7) pp 3670-3681 DOI: 10.1016/j.eswa.2014.12.013
- [7] Chen B Chen L Chen Y 2013 Efficient ant colony optimization for image feature selection *Signal Processing* vol 93(6) pp 1566-1576 DOI: 10.1016/j.sigpro.2012.10.022
- [8] Liang L et al. 2016 Feature selection for machine fault diagnosis using clustering of non-negation matrix factorization *Measurement* vol 94 pp 295-305 DOI: 10.1016/j.measurement.2016.08.003

- [9] Liang D Tsai C F Wu H T 2015 The effect of feature selection on financial distress prediction *Knowledge-Based Systems* vol 73 pp 289-297 DOI: 10.1016/j.knosys.2014.10.010
- [10] Caballero E Molina J 2003 Programación Multiobjetivo mediante un procedimiento de búsqueda tabú *Actas del II Congreso Español de Metaheurísticas y Algoritmos Evolutivos y Bioinspirados*
- [11] Caballero R Gandibleux X Molina J 2004 MOAMP: A generic multiobjective metaheuristic using an adaptive memory *Technical report*, University of Valenciennes, Valenciennes, France
- [12] Glover F Laguna M 2013 Tabu Search *Handbook of Combinatorial Optimization* Du D Z Pardalos P M (Boston: Springer) pp 3261-3362 DOI: 10.1007/978-1-4613-0303-9_33
- [13] Glover F Laguna M Marti R 2007 Principles of tabu search *Handbook of Approximation algorithms and metaheuristics* Gonzalez T (Boca Raton: Chapman & Hall/CRC) vol 23 pp 1-12
- [14] Gómez J Pacheco J Gonzalo-Orden H 2015 A Tabu Search Method for a Bi-Objective Urban Waste Collection Problem *ComputerAided Civil and Infrastructure Engineering* vol 30(1) pp 36-53 DOI: 10.1111/mice.12031
- [15] Martínez-Puras A Pacheco J 2016 MOAMP-Tabu search and NSGA-II for a real Bi-objective scheduling-routing problem *Knowledge-Based Systems* vol 112 pp 92-104 DOI: 10.1016/j.knosys.2016.09.001
- [16] Pacheco J et al. 2013 Bi-objective bus routing: an application to school buses in rural areas *Transportation Science* vol 47(3) pp 397-411 DOI: 10.1287/trsc.1120.0437
- [17] Cardona-Valdés Y Álvarez A Pacheco J 2014 Metaheuristic procedure for a bi-objective supply chain design problem with uncertainty *Transportation Research Part B: Methodological* vol 60 pp 66-84 DOI: 10.1016/j.trb.2013.11.010
- [18] Caballero R et al. 2007 Solving a multiobjective location routing problem with a metaheuristic based on tabu search. Application to a real case in Andalusia *European Journal of Operational Research* vol 177(3) pp 1751-1763 DOI: 10.1016/j.ejor.2005.10.017
- [19] Martí R et al. 2015 Multiobjective GRASP with path relinking *European Journal of Operational Research* vol 240(1) pp 54-71 DOI: 10.1016/j.ejor.2014.06.042
- [20] Zitzler E Thiele L 1998 Multiobjective optimization using evolutionary algorithms - A comparative case study Parallel Problem Solving from Nature Lecture Notes in Computer Science Eiben A E Bäck T Schoenauer M Schwefel H P (Berlin: Springer) vol 1498 pp 292-301