Multiple Reference Points MOEA/D for Feature Selection

Hoai Bach Nguyen, Bing Xue
School of Engineering and Computer Scince
Victoria University of Wellington
Wellington, New Zealand
{Hoai.Bach.Nguyen,Bing.Xue}@ecs.vuw.ac.nz

Hisao Ishibuchi
Department of Information Science and Engineering
Southern University of Science and Technology
Shenzhen, China
Hisaoi@cs.osakafu-u.ac.jp

Peter Andreae, Mengjie Zhang
School of Engineering and Computer Scince
Victoria University of Wellington
Wellington, New Zealand
{Peter.Andreae,Mengjie.Zhang}@ecs.vuw.ac.nz

ABSTRACT

Feature selection can be considered a multi-objective problem since its two main objectives usually conflict with each other. Many Pareto dominance-based algorithms have been applied to feature selection. However, feature subsets evolved by these algorithms are mostly around the center of the Pareto front. MOEA/D can avoid this issue to some extent, but still needs to be modified when applying it to solve complex feature selection problems. This paper proposes a new decomposition strategy for feature selection called MOEA/D-MRPs which uses multiple reference points instead of multiple weight vectors. The proposed algorithm, is evaluated on eight different datasets and compared with three Pareto dominance-based algorithms and the standard MOEA/D algorithm. Experimental results show that MOEA/D-MRPs can efficiently evolve a more diverse set of non-dominated solutions than three Pareto dominance-based algorithms and achieve better classification performance than the standard MOEA/D algorithm. On large datasets, MOEA/D-MRPs is also the most efficient algorithm.

CCS CONCEPTS

•**Computing methodologies → Feature selection;**

KEYWORDS

Multi-objective, Feature Selection, MOEA/D

ACM Reference format:
Hoai Bach Nguyen, Bing Xue, Hisao Ishibuchi, and Peter Andreae, Mengjie Zhang. 2017. Multiple Reference Points MOEA/D for Feature Selection. In *Proceedings of GECCO '17 Companion, Berlin, Germany, July 15-19, 2017,* 2 pages.
DOI: http://dx.doi.org/10.1145/3067695.3075985

1 INTRODUCTION

A rapid growth in technologies results in datasets with a large number of features, which often causes an intensive computation cost and a poor performance in a classification system. Feature selection aims to select a small proportion of the original features ($fRate$) and reduce the classification error (*eRate*). These two main objectives are usually in conflict, so feature selection can be considered as a multi-objective problem.

Evolutionary Computation (EC) is a family of population-based optimization algorithms, which provide a set of non-dominated solutions to a multi-objective problem in a natural way. Several evolutionary multi-objective (EMO) algorithms evaluate candidate solutions by using a Pareto dominance relation. Although Pareto dominance-based algorithms usually work well on problems having two or three objectives, they can not find optimal or near optimal solutions on the edges of the Pareto front for combinatorial problems [3] such as knapsack or feature selection [7]. MOEA/D [4], a representative of scalarising function-based algorithms, is usually better than Pareto dominance-based algorithms in terms of preserving diversity and efficiency [2]. MOEA/D decomposes a multi-objective problem to many scalar sub-problems by a set of weight vectors. Initializing the weight vectors is a challenging task since it depends on the true Pareto front, which is unknown in most real-world problems. In this work, we will propose a novel decomposition method for MOEA/D to solve feature selection problems, which can provide an even distribution of solutions regardless of the true Pareto front shape.

2 A NOVEL DECOMPOSITION STRATEGY

In the standard MOEA/D framework, a multi-objective problem is decomposed by a set of weight vectors. In our work, a set of reference points, uniformly allocated on the $fRate$ axis, is used instead of weight vectors to guide the search for solutions on the Pareto front. Each reference point represents the ideal feature subset with 0% classification error and ($n_{ref} = refRate*n$) features, where n is the total number of original features. A reference point defines a sub-problem of finding an optimal solution on the Pareto front with the same or smaller number of features. The search spaces of the sub-problems are smaller than the original search space. The task of MOEA/D is to push each candidate solution of a sub-problem towards the reference point, so that it can find the optimal solution. The fitness function of each sub-problem is designed as follows.

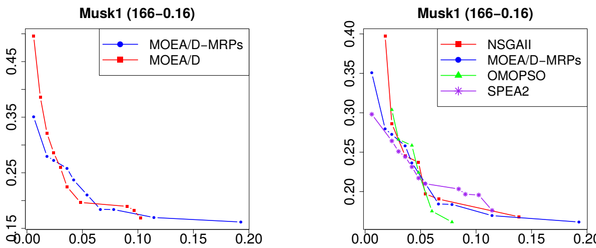
$$fitness(S) = eRate(S) + 100 * max(|S| - n_{ref}, 0) + 0.01 * |S| \quad (1)$$

where S is a feature subset which contains $|S|$ features.

The primary component of the fitness function is the *eRate(S)*, which must be minimized. The second component is a penalty value to ensure that the number of selected features in S should not exceed n_{ref}. The last component shows a weak preference

GECCO '17 Companion, Berlin, Germany

DOI: http://dx.doi.org/10.1145/3067695.3075985



(a) MOEA/D-MRPs vs standard MOEA/D

(b) MOEA/D-MRPs vs Pareto dominance algorithms

Figure 1: Best Pareto fronts

for a smaller feature subset among feature subsets with the same classification error but different numbers of features.

MOEA/D uses neighborhood to guide its search. The neighbors of a sub-problem are sub-problems, whose reference points are close to the sub-problem's reference point.

When MOEA/D constructs a new candidate for a sub-problem with constraints, it may have to repair the candidate. In the multiple reference points strategy, each feature subset of a sub-problem should contain at most n_{ref} features. To repair infeasible feature subsets, a simple strategy is to remove features until the constraint is satisfied. Firstly, each selected feature of a sub-problem will be used to classify the training dataset to obtain its classification accuracy. The repair process starts from the selected feature with the lowest classification performance. One might argue that the removed feature may work well with the remaining features despite the low classification performance by itself. In this case, the neighbor sub-problem with a slightly larger n_{ref} will contain the removed feature and this information is sent to the sub-problem in the next step of the search and guides it to re-select the removed feature. Since all classification accuracies of single features are calculated in advance, this removing process is very efficient. The proposed algorithm is called MOEA/D-MRPs.

3 EXPERIMENT DESIGN

MOEA/D-MRPs was compared to the standard MOEA/D and the three Pareto dominance-based algorithms (NSGA-II [1], SPEA2 [8] and OMOPSO [6]) on eight datasets selected from the UCI machine learning repository [5]. The candidate solutions are evaluated by K-nearest neighbor (KNN) where K is set to 5. For all algorithms, the population size and the maximum number of iterations are 100. In each run, the datasets are divided into training and test sets with the proportions of 70% and 30%, respectively. During the training process, KNN with 10-fold cross-validation is applied to calculate the classification error rate. The evolved feature subsets are then evaluated on the test set to obtain their testing accuracies.

4 RESULTS AND ANALYSIS

Each algorithm is run 50 independent times resulting in 50 sets of non-dominated solutions, which are combined into a single set of solutions. From the union set, all the non-dominated solutions are selected to form the best Pareto front, as shown in Figure 1. The figures represent the best Pareto fronts on the test sets of the Musk1 dataset. In each sub-figure, the two numbers inside the brackets show the total number of original features and the testing accuracy

Table 1: Hypervolume Results

MRPs	MOEA/D	NSGA-II	SPEA2	OMOPSO
0.774	0.753(+)	0.739(+)	0.759(+)	0.723(+)

of using all features. The horizontal and vertical axes represent $fRate$ and $eRate$, respectively. As can be seen from the figure, despite selecting less than 20% original features, MOEA/D-MRPs can evolve at least one feature subset that is better than the set of all features on both training and test sets. Although the patterns of Pareto fronts evolved by MOEA/D-MRPs and standard MOEA/D are quite similar, given the same number of features MOEA/D-MRPs usually achieves lower error rate. In comparison with the three Pareto-dominance algorithms, the non-dominated solutions evolved by MOEA/D-MRPs are distributed more evenly on the objective space.

The hypervolume indicator is used to examine the five algorithms. Since each algorithm is run 50 independent times resulting in 50 hypervolume values, a Wilcoxon test with significance level of 0.05 is used to compare between MOEA/D-MRPs and other algorithms. As can be seen from Table 1, the proposed algorithm achieves a significantly higher hypervolume value than the other algorithms, which shows that MOEA/D-MRPs can evolve better Pareto fronts. "=", "+", "-" mean that MOEA/D-MRPs is similar, significantly better or worse than the other algorithms, respectively.

5 CONCLUSIONS

This paper proposed a new decomposition strategy for the multi-objective feature selection problem where a set of evenly distributed reference points on the $fRate$ axis is used instead of a weight vector set. Experimental results show that MOEA/D-MRPs achieves better classification accuracy than standard MOEA/D and evolves more evenly distributed non-dominated solutions than the three Pareto dominance-based algorithms. By allocating individuals evenly with respect to the feature rate, MOEA/D-MRPs performs more efficiently than other algorithms on large datasets.

REFERENCES

[1] Kalyanmoy Deb, Samir Agrawal, Amrit Pratap, and T. Meyarivan. 2000. *A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II*. Springer Berlin Heidelberg, Berlin, Heidelberg, 849–858.
[2] Hisao Ishibuchi, Yuji Sakane, Noritaka Tsukamoto, and Yusuke Nojima. 2009. Adaptation of scalarizing functions in MOEA/D: An adaptive scalarizing function-based multiobjective evolutionary algorithm. In *International Conference on Evolutionary Multi-Criterion Optimization*. Springer, 438–452.
[3] Andrzej Jaszkiewicz. 2004. On the computational efficiency of multiple objective metaheuristics. The knapsack problem case study. *European Journal of Operational Research* 158, 2 (2004), 418–433.
[4] Hui Li and Qingfu Zhang. 2009. Multiobjective optimization problems with complicated Pareto sets, MOEA/D and NSGA-II. *IEEE Transactions on Evolutionary Computation* 13, 2 (2009), 284–302.
[5] M. Lichman. 2013. UCI Machine Learning Repository Irvine, CA: University of California, School of Information and Computer Sciences. (2013). http://archive.ics.uci.edu/ml
[6] Margarita Reyes Sierra and Carlos A. Coello Coello. 2005. *Improving PSO-Based Multi-objective Optimization Using Crowding, Mutation and e-Dominance*. Springer, 505–519.
[7] Bing Xue, Mengjie Zhang, and Will N Browne. 2013. Particle swarm optimization for feature selection in classification: a multi-objective approach. *IEEE transactions on cybernetics* 43, 6 (2013), 1656–1671.
[8] Eckart Zitzler, Marco Laumanns, Lothar Thiele, and others. 2001. SPEA2: Improving the strength Pareto evolutionary algorithm. In *Eurogen*, Vol. 3242. 95–100.