Information Sciences 465 (2018) 388–403

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Preselection via classification: A case study on evolutionary multiobjective optimization

Jinyuan Zhang[a], Aimin Zhou[a,*], Ke Tang[b], Guixu Zhang[a]

[a] *Shanghai Key Laboratory of Multidimensional Information Processing, Department of Computer Science and Technology, East China Normal University, Shanghai 200062, China*

[b] *Shenzhen Key Laboratory of Computational Intelligence, Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, Guangdong 518055, China*

ARTICLE INFO

Article history:
Received 5 January 2018
Revised 22 June 2018
Accepted 24 June 2018
Available online 12 July 2018

Keywords:
Preselection
Classification
Multiobjective optimization
Evolutionary algorithm

ABSTRACT

In evolutionary algorithms, a preselection operator aims to select the promising offspring solutions from a set of candidate offspring solutions. It is usually based on the estimated or real objective values of the candidate offspring solutions. In a sense, the preselection can be treated as a classification procedure, which classifies the candidate offspring solutions into promising ones and unpromising ones. Following this idea, in this paper we propose a *classification based preselection (CPS)* strategy for evolutionary multiobjective optimization. When applying CPS, an evolutionary algorithm maintains two external populations (training data set) that consist of some selected 'good' and 'bad' solutions; then it trains a classifier based on the training data set in each generation. Finally, it uses the classifier to filter the unpromising candidate offspring solutions and choose a promising one from the generated candidate offspring set for each parent solution. In such cases, it is not necessary to evaluate or estimate the objective values of the candidate offspring solutions. In this study, CPS is applied to three state-of-the-art *multiobjective evolutionary algorithms (MOEAs)* and is empirically studied on two sets of test instances. The results suggest that CPS can improve the performance of these MOEAs.

1. Introduction

In *evolutionary algorithms (EAs)*, the preselection operator is a component that has different meanings [24]. In this paper, we use the term 'preselection' to denote the process that selects the promising offspring solutions from a set of candidate offspring solutions created by the offspring generation operators before the environmental selection procedure.

A key issue in preselection is how to measure the quality of the candidate offspring solutions. The surrogate model (metamodel) is one of the major techniques for this [14,16,32,36]. The surrogate model is a method to mimic the original optimization surface by finding an alternative mapping, which is computationally cheaper, from the decision variable (model input) to the dependent variable (model output) through a given training data set. Some popular surrogate models include the Gaussian process [8,10,11,21,42], radial basis function [31], artificial neural networks [15], and polynomial response surfaces [34]. In the community of evolutionary computation, the surrogate model is usually used to replace the original

* Corresponding author. Tel.: +862162233560.
E-mail address: amzhou@cs.ecnu.edu.cn (A. Zhou).

https://doi.org/10.1016/j.ins.2018.06.073
0020-0255/

objective function, especially when the objective function evaluation is expensive [14,22,23,32,35]. In the case of preselection, surrogate model based preselection strategies have been employed to solve different optimization problems [10,12,16,26].

In EAs, the preselection can be naturally regarded as a classification problem. More precisely, the preselection classifies the candidate offspring solutions into two categories: the selected promising ones, and the discarded unpromising ones. This indicates that what we need to know is whether a candidate offspring solution is good or bad instead of how good it is. Following this idea, a classification method was proposed for the expensive optimization problem [23]. This work has also been extended by using both classification and regression methods in preselection [22]. A major difference between the regression model based approach and the classification based approach is that the former measures the quality of the candidate offspring solutions precisely while the latter roughly classifies the candidate offspring solutions into several categories. For a candidate offspring solution, an accurate quality measurement might be useful; however, in (pre)selection, a label has already offered enough information for making decisions.

It might be more suitable to use classification techniques in multiobjective optimization. The reason is that the solutions in a *multiobjective evolutionary algorithm (MOEA)* are either dominated or nondominated, which forms two classes naturally. In [4], the classification algorithms are used to learn Pareto dominance relations. As far as we know, we are the first ones to apply classification to the preselection of MOEAs [39,40]. Very recently, a similar idea has been implemented in [20].

In this paper, we extend our previous work [39,40] and propose a general *classification based preselection (CPS)* scheme for evolutionary multiobjective optimization. The major procedure of CPS works as follows: in each generation, firstly a training data set (external populations) is updated by recently found solutions; secondly, a classifier is built according to the training data set; thirdly, for each solution, candidate offspring solutions are generated and their labels are predicted by the classifier; finally for each solution, an offspring solution is chosen according to the predicted labels. The major differences between this paper and the previous work are as follows.

- The data preparation strategy is improved: all the dominated solutions found so far are used to update the negative training set, and the training set contains more points. In [39,40], only the solutions in the previous generation are used to update the negative training set.
- A general CPS strategy is proposed and applied to the three main MOEA frameworks: the Pareto domination based MOEA, the indicator based MOEA, and the decomposition based MOEA. In [39,40], CPS is applied to one framework.
- A systematic empirical study, based on two sets of test instances, has been performed to demonstrate the advantages of CPS.

The rest of the paper is organized as follows. Section 2 presents the related work on evolutionary multiobjective optimization and briefly introduces the main MOEA frameworks used in the paper. Section 3 presents the CPS in detail. The proposed CPS is systematically studied on some benchmark problems in Sections 4 and 5. Finally, Section 6 concludes this paper with some future work remarks.

2. Evolutionary multiobjective optimization

In this paper, we consider box-constrained continuous *multiobjective optimization problems (MOPs)* which can be mathematically formulated as follows;

$$\begin{aligned} \min \quad & F(x) = (f_1(x), \cdots, f_m(x))^T \\ \text{s.t} \quad & x \in \Pi_{i=1}^{n}[a_i, b_i] \end{aligned} \tag{1}$$

where $x = (x_1, \cdots, x_n)^T \in R^n$ is a decision variable vector; $\Pi_{i=1}^{n}[a_i, b_i] \subset R^n$ defines the feasible region of the search space; $a_i < b_i$ $(i = 1, \cdots, n)$ are the lower and upper boundaries of the search space respectively, $f_j : R^n \rightarrow R (j = 1, \cdots, m)$ is a continuous mapping, and $F(x)$ is an objective vector.

Due to the conflicting nature among the objectives in (1), there usually does not exist a single solution that is able to optimize all the objectives at the same time. Therefore, a set of tradeoff solutions, named as *Pareto optimal solutions*, are of interest. The set of all the Pareto optimal solutions is called the *Pareto set (PS)* in the decision space and the *Pareto front (PF)* in the objective space.

In real world applications, a practical target of multiobjective optimization algorithms is to find an approximation to the PF (PS) of (1). Among different methods, EAs have become useful for achieving this target due to their population based search property, which makes an MOEA be able to approximate the PF (PS) in a single run. A large number of MOEAs have been proposed in the last few years, and most of these algorithms can be roughly partitioned into three classes: the Pareto domination based MOEAs [7,9,13,29,37,46,47], the indicator based MOEAs [1–3,5,17,27,45], and the decomposition based MOEAs [18,25,28,33,38,41]. In this paper, we demonstrate that CPS is able to improve the performance of the three kinds of MOEAs. Three algorithms- the *regularity model based multiobjective estimation of distribution algorithm (RM-MEDA)* [43], the *hypervolume metric selection based EMOA (SMS-EMOA)* [5], and the *MOEA based on decomposition with multiple differential evolution mutation operators (MOEA/D-MO)* [19]- from the three MOEA categories, respectively, are used as the basic algorithms. These algorithms are briefly introduced here, and more details are available in the corresponding papers.