Computers & Operations Research 53 (2015) 128–153

Contents lists available at ScienceDirect

Computers & Operations Research

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

Constraint-handling through multi-objective optimization: The hydrophobic-polar model for protein structure prediction

Mario Garza-Fabre *, Eduardo Rodriguez-Tello, Gregorio Toscano-Pulido

Information Technology Laboratory, CINVESTAV-Tamaulipas, Parque Científico y Tecnológico TECNOTAM, Km. 5.5 carretera Cd. Victoria-Soto La Marina, Cd. Victoria, Tamaulipas 87130, Mexico

ARTICLE INFO

Available online 7 August 2014

Keywords:
Constraint-handling
Evolutionary multi-objective optimization
Fitness landscape analysis
Search bias
Protein structure prediction
Hydrophobic-polar model

ABSTRACT

In the multi-objective approach to constraint-handling, a constrained problem is transformed into an unconstrained one by defining additional optimization criteria to account for the problem constraints. In this paper, this approach is explored in the context of the hydrophobic-polar model, a simplified yet challenging representation of the protein structure prediction problem. Although focused on such a particular case of study, this research work is intended to contribute to the general understanding of the multi-objective constraint-handling strategy. First, a detailed analysis was conducted to investigate the extent to which this strategy impacts on the characteristics of the fitness landscape. As a result, it was found that an important fraction of the infeasibility translates into neutrality. This neutrality defines potentially shorter paths to move through the landscape, which can also be exploited to escape from local optima. By studying different mechanisms, the second part of this work highlights the relevance of introducing a proper search bias when handling constraints by multi-objective optimization. Finally, the suitability of the multi-objective approach was further evaluated in terms of its ability to effectively guide the search process. This strategy significantly improved the performance of the considered search algorithms when compared with respect to commonly adopted techniques from the literature.

1. Introduction

Evolutionary computation methods and other metaheuristic algorithms have been successfully used to solve complex optimization problems which arise in a diversity of scientific and engineering applications. Often, however, optimization involves not only to reach the best value for a given objective function (or set of objective functions), but also to satisfy a certain set of predefined requirements called constraints. Therefore, additional mechanisms need to be implemented within metaheuristic algorithms in order to search effectively through this kind of constrained solution spaces.

The *hydrophobic-polar* (HP) model [1,2] is an abstract formulation of the *protein structure prediction* (PSP) problem, where hydrophobicity is assumed to be the main stabilizing force in the protein folding process. Under this model, PSP is defined as the problem of finding a *self-avoiding embedding* of the protein chain on a given lattice, such that the interaction among hydrophobic amino acids is maximized. From the computational point of view, the HP model entails a challenging problem in combinatorial optimization [3,4]. One of the main sources of difficulty in this problem lies in the fact that, using the existing problem representations, a significant portion of the solution space encodes infeasible (non-self-avoiding) protein structures. Hence, it is important to devise effective mechanisms for handling the constraints that this problem presents. Two main research directions have been adopted to cope with this issue. On the one hand, the search can be confined to the space of only feasible, self-avoiding protein conformations. On the other hand, infeasible protein conformations can also be taken into consideration, which has been achieved in the literature by implementing a *penalty strategy*. From the literature, however, it is not possible to identify a clear consensus on which of the two directions, *i.e.*, to avoid or to consider infeasible conformations, could lead to the development of more efficient metaheuristics for solving this problem [5–9].

Premised upon the belief that infeasible conformations can provide valuable information for guiding the search process, this

* Corresponding author. Tel.: +52 834 107 0220.
E-mail addresses: mgarza@tamps.cinvestav.mx (M. Garza-Fabre), ertello@tamps.cinvestav.mx (E. Rodriguez-Tello), gtoscano@tamps.cinvestav.mx (G. Toscano-Pulido).

http://dx.doi.org/10.1016/j.cor.2014.07.010
0305-0548/

research work inquires into the use of multi-objective optimization as an alternative constraint-handling strategy for the HP model. Particularly, constraints in the HP model are treated as a supplementary optimization criterion, leading to an unconstrained multi-objective problem.[1] Using such an alternative formulation of the HP model, infeasible solutions can become incomparable with respect to feasible ones, having thus better opportunities for participating throughout the search process. In contrast to the penalty strategy, which represents one of the most widely used techniques in the constraint-handling literature, in essence the multi-objective (MO) method does not require the fine-tuning of the penalty parameters[2]; in the penalty strategy, finding the right balance between objective function and penalty values has been regarded to be a difficult optimization problem itself [10,11]. The use of multi-objective optimization for handling constraints is not a novel idea; recent reviews on this topic can be found in [11,12]. Nevertheless, it was not until recently that the preliminary results of this research reported for the first time, to the best of the authors' knowledge, the application of the MO constraint-handling strategy to the particular HP model of the PSP problem [13].

Building further on this research, the primary aim of this study is to contribute to the general understanding of the functioning of the MO constraint-handling technique. First, a detailed analysis is conducted in order to investigate the potential effects of the problem transformation from the perspective of the fitness landscape. More specifically, it is evaluated how the use of the MO problem formulation impacts on an important property of the fitness landscape: *neutrality*. It has been argued that the MO approach to constraint-handling could be rather ineffective if a search bias towards the feasible region is not introduced [14]. Therefore, the second part of this document concerns the study of different mechanisms which can be employed for providing the MO strategy with such a search bias. The last part of this research work extends the comparative analysis reported in [13], where the MO approach is evaluated with respect to commonly adopted techniques from the specialized literature. While the preliminary results presented in [13] assumed a fixed biasing scheme for the MO method and focused only on the performance of a population-based algorithm, the different biasing mechanisms analyzed in the second part of this study, as well as both single-solution-based and population-based algorithms, have been included in the present study. Likewise, only 15 test instances for the two-dimensional HP model (based on the square lattice) were used in [13]. In contrast, the present study covers also the three-dimensional case (based on the cubic lattice) and a total of 30 test cases have been considered.

The remainder of this document is organized as follows. Section 2 provides background concepts and sets the notation used in this study. Section 3 reviews related work on constraint-handling methods for the HP model as well as on the topic of single-objective to multi-objective transformations. The studied MO constraint-handling approach is described in Section 4. Section 5 presents the analysis with regard to the fitness landscape transformation. The search bias issue is addressed in Section 6. The comparative study which focuses on search performance is covered in Section 7. Finally, Section 8 discusses the main findings and presents the conclusions of this study. Appendices at the end of this document contain supplementary information with regard to implementation details of the considered search algorithms, performance measures, test instances, the methodology followed for the statistical significance analyses, and the utilized experimental platform.

2. Background concepts and notation

2.1. Single-objective and multi-objective optimization

Without loss of generality, a *single-objective optimization problem* can be formally stated as follows:

$$\begin{aligned}&\text{Minimize} \quad f(\mathbf{x}),\\ &\text{subject to} \quad \mathbf{x} \in \mathcal{X}_{\mathcal{F}},\end{aligned} \tag{1}$$

where $\mathbf{x}$ is a *solution vector*; $\mathcal{X}_{\mathcal{F}}$ denotes the *feasible set*, *i.e.*, the set of all *feasible solution vectors* in the search space $\mathcal{X}$, $\mathcal{X}_{\mathcal{F}} \subsetneq \mathcal{X}$; and $f : \mathcal{X} \to \mathbb{R}$ is the objective function to be optimized. The aim is thus to find the feasible solution(s) yielding the optimum value for the objective function; that is, to find $\mathbf{x}^* \in \mathcal{X}_{\mathcal{F}}$ such that $f(\mathbf{x}^*) = \min\{f(\mathbf{x}) | \mathbf{x} \in \mathcal{X}_{\mathcal{F}}\}$.

Similarly, a *multi-objective optimization problem* can be formally defined as follows:

$$\begin{aligned}&\text{Minimize} \quad \mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_k(\mathbf{x})]^T,\\ &\text{subject to} \quad \mathbf{x} \in \mathcal{X}_{\mathcal{F}},\end{aligned} \tag{2}$$

where $\mathbf{f}(\mathbf{x})$ is the *objective vector* and $f_i : \mathcal{X} \to \mathbb{R}$ is the ith objective function, $i \in \{1, 2, \ldots, k\}$. Rather than searching for a single optimal solution, the task in multi-objective optimization is to identify a set of trade-offs among the conflicting objectives. More formally, the goal is to find a set of *Pareto-optimal solutions* $\mathcal{P}^*$, such that $\mathcal{P}^* = \{\mathbf{x}^* \in \mathcal{X}_{\mathcal{F}} | \nexists \mathbf{x} \in \mathcal{X}_{\mathcal{F}} : \mathbf{x} \prec \mathbf{x}^*\}$. The symbol "$\prec$" denotes the *Pareto-dominance* relation [15]:

$$\begin{aligned}&\mathbf{x} \prec \mathbf{x}' \Leftrightarrow \forall i \in \{1, \ldots, k\} : f_i(\mathbf{x}) \le f_i(\mathbf{x}') \wedge\\ &\exists j \in \{1, \ldots, k\} : f_j(\mathbf{x}) < f_j(\mathbf{x}').\end{aligned} \tag{3}$$

If $\mathbf{x} \prec \mathbf{x}'$, then $\mathbf{x}$ is said to *dominate* $\mathbf{x}'$. Otherwise, $\mathbf{x}'$ is said to be *nondominated* with respect to $\mathbf{x}$, denoted by $\mathbf{x} \nprec \mathbf{x}'$. The image of $\mathcal{P}^*$ in the objective space is the so-called *Pareto-optimal front*, usually also referred to as the *trade-off surface*.

2.2. Fitness landscapes and neutrality

The notion of a fitness landscape, first introduced by Wright [16], has been found to be useful in understanding the most essential characteristics of certain optimization problems, or problem classes. By analyzing the fitness landscape, it is possible to gain further insight into problem difficulty as a means of explaining, or even predicting, the performance of search algorithms. Fitness landscape analysis is expected to provide important clues for guiding the development of more competitive search mechanisms, which are able to deal with (or to take advantage of) the particular characteristics of the given optimization task. Some fundamental definitions on this topic, which are relevant according to the scope of this study, are presented below. For a more comprehensive literature review on fitness landscapes analysis the reader can be referred to [17–21].

A *fitness landscape* can be generally defined in terms of a triplet $(\mathcal{X}, \mathcal{N}, \xi)$. The first element, $\mathcal{X}$, represents the set of all potential solutions to the problem, *i.e.*, the *search space*. The notion of connectedness among solutions in $\mathcal{X}$ is introduced by the so-called *neighborhood structure*, $\mathcal{N} : \mathcal{X} \to 2^{\mathcal{X}}$, a function which maps each possible solution $\mathbf{x} \in \mathcal{X}$ to a set of solutions $\mathcal{N}(\mathbf{x}) \subseteq \mathcal{X}$. Hence, $\mathcal{N}(\mathbf{x})$ is referred to as the neighborhood of $\mathbf{x}$ and each $\mathbf{x}' \in \mathcal{N}(\mathbf{x})$ is called a *neighbor* of $\mathbf{x}$. Finally, ξ denotes the evaluation scheme, consisting of (i) a measure (or set of measures) to serve as an indicator of the quality of the different solution candidates; and (ii) a mechanism to impose an ordering relation

[1] The process of restating a single-objective problem as a multi-objective one is usually referred to as *multi-objectivization*; refer to Section 3.2.

[2] However, the MO strategy may require additional parameters or the combination with other mechanisms for biasing purposes.