General Aspect-based Selection Concept for Multi- and Many-Objective Molecular Optimization

Susanne Rosenthal[1]
[1]Steinbeis Innovation Center "Intelligent and Self-Optimizing Software Assistance Systems"
Paracelsusstr. 9
Bergisch Gladbach, Germany
Susanne.Rosenthal@stw.de

Markus Borschbach[1,2]
[2]University of Applied Sciences, FHDW
Hauptstr. 2
Bergisch Gladbach, Germany
Markus.Borschbach@{fhdw.de,stw.de}

ABSTRACT

Many-objective optimization is steadily gaining importance in the field of automated drug design involving simultaneous optimization of several physiochemical and biological properties. Pareto-based MOEAs slow down in convergence as the classification of the solutions' quality according to the Pareto dominance principle becomes increasingly undifferentiated with the rise of objectives. This research addresses the enhancement of a MOEA for drug design with the aim of solving many-objective molecular optimization problems. For this purpose, a sophisticated selection concept is designed. This selection strategy is front-based, but the Pareto dominance principle is applied to a two-dimensional indicator problem and not directly to the optimization problem. The first indicator reflects the solutions' quality with regard to the objective values and the second indicator refers to the general aspect in molecular optimization - the genetic dissimilarity among the solutions in a population. First experiments reveal that this selection strategy is able to identify a selected number of improved molecules within 10 generations for a 3D- and 4D-molecular optimization problem.

CCS CONCEPTS

•Theory of computing → Design and analysis of algorithm; •Mathematics of computing → Bio-inspired Optimization; *Redundancy;* Evolutionary Algorithm;

KEYWORDS

many-objective molecular optimization, selection strategy

ACM Reference format:
Susanne Rosenthal[1] and Markus Borschbach[1,2]. 2017. General Aspect-based Selection Concept for Multi- and Many-Objective Molecular Optimization. In *Proceedings of GECCO '17 Companion, Berlin, Germany, July 15-19, 2017,* 2 pages.
DOI: http://dx.doi.org/10.1145/3067695.3082046

1 INTRODUCTION

Many-objective Optimization Problems (MaOPs) are defined as Multi-objective Optimization Problems (MOPs) with more than three objectives. MaOPs arise out of real world applications and pose a challenge for MOEAs in both targets, convergence and diversity. Pareto-based MOEAs have difficulties to solve MaOPs due to their inability to classify the quality of solutions by the Pareto dominance principle. Furthermore, the definition of diversity is less straightforward to reformulate in MaOPs.[4]

Different approaches of Many-objective Evolutionary Algorithms (MaOEAs) have been evolved in the past addressing the challenge of convergence and diversity by methods of objective reduction, incorporation and preferences, modified dominance definitions and the introduction of additional selection criteria. An overview of these algorithms is given in [1].

A MOEA for molecular optimization, referred to as COSEA-MO, has been recently reported in [3] identifying a selected number of highly qualified molecules within a very low number of generations. COSEA-MO is evolved to complement an *in vitro* drug design process as a computer-assisted system to identify a selected number of improved molecules providing a wide range of genetic diversity within a very low iteration number for an efficient laboratory examination. Dynamic deterministic variation operators are used in COSEA-MO and a mating pool of the old population and the offspring is generated after variation. A combination of fitness-proportionate and indicator-based selection determines the individuals of the succeeding generation. As the selection strategy is partly Pareto front-based, a more sophisticated selection strategy is required, which is target-oriented to many-objective molecular optimization. The selection strategy presented in the following applies the Pareto dominance not directly to the optimization problem, but to a two-dimensional indicator problem covering the two generic aspects of molecular optimization problems: firstly, an indicator for the quality of the molecules; secondly, an indicator for the genetic dissimilarity of a molecule with regard to the current population. This selection concept is introduced in the following and the performance of COSEA-MO with the traditional and the sophisticated selection strategy, further termed nCOSEA-MO, are compared for a 3D- and 4D-molecular optimization problem as presented in [3].

2 CONCEPT OF SELECTION STRATEGY

A MaOP is given by $f : P \longrightarrow \mathbb{R}^m,\ p \longrightarrow (f_1(p), f_2(p), ..., f_m(p))$, whereby $m > 3$ is the number of objective molecular functions f_i which have to be minimized, and P is the quantity of feasible molecules. The procedure of the novel selection strategy is described in Algorithm1. The strategy is ranked and binary tournament based. The Pareto principle used for ranking is not directly

GECCO '17 Companion, Berlin, Germany
 978-1-4503-4939-0/17/07…$15.00
DOI: http://dx.doi.org/10.1145/3067695.3082046

applied on the objective values but on a two-dimensional indicator-based minimization problem (line 4). The first indicator reflects the solutions' quality by the calculation of the L_p-norm of the objective values to a reference point (line 2), which is determined by the minimum of each objective provided by the population members (line 1). Therefore, this reference point varies with the population. The second indicator refers to the general aspect of maintaining a high genetic dissimilarity within the populations. Needleman Wunsch Algorithm (NMW) [2] is chosen as global sequence alignment (line 3). The N-best individuals are selected in the succeeding

Algorithm 1: Pseudo code of the selection strategy

Input: Current population P_t with $|P_t| = 2N$, $P_{t+1} = \{\}$
Calculation of the two indicator values for each solution:
1: $f_{min} := (min_{i_1} f_1(p_{i_1}), min_{i_2} f_2(p_{i_2}), ..., min_{i_m} f_m(p_{i_m}))$;
2: $\forall p \in P_t$: $f_{L_{p-norm}}(p) = L_p(f(p), f_{min})$;
3: $\forall p \in P_t$: $diss(p) = \frac{1}{|P_t|} \sum_{p \in P_t} SequenceAlignment(p, P_t - p)$;
Selection process:
4: Ranking of P_t according to $(f_{L_p-norm}, diss)$ into fronts F_i;
5: **while** $|P_{t+1}| + |F_i| < N$ **do**
  $P_{t+1} = P_{t+1} \cup F_i$; i++;
end
6: binary tournament selection: **while** $|P_{t+1}| < N$ **do**
  select $p_1, p_2 \in P_t \setminus \{P_{t+1}\}$:
  if $(f_{L_p}(p_1) * diss(p_1) < f_{L_p}(p_2) * diss(p_2))$ add p_1 to P_{t+1} ;
  else add p_2 to P_{t+1};
end

generation based on the rank (line 5) and the volume dominance principle via binary tournament selection (line 6).

3 EXPERIMENTS

The performances of COSEA-MO and nCOSEA-MO are compared for a 3D- and 4D-molecular minimization problem [3]. Molecular weight, average hydrophilicity and NMW as a similarity score to a predefined reference peptide are the 3D-MOP. The addition of the Instability Index to the 3D-MOP becomes the 4D-MaOP. The test runs are performed with a population size of 100, 20-mer peptides as individuals, L_2-norm and 10 generations repeated for 30 times. The approximate Pareto optimal sets (PFs) of COSEA-MO and nCOSEA-MO in each generation are compared in terms of the established C-metric: $C(PF_1, PF_2) := |\{b \in PF_2 \mid \exists a \in PF_1 : a \preceq b\}| / |PF_2|$. PF of COSEA-MO is determined according to the molecular optimization problem, whereas PF of COSEA-MO is determined according to the two-dimensional indicator problem. The C-metric values are determined according to the objective values as usual. Table 1 and 2 depict the C-metric values $C_1 = C$(COSEA-MO, nCOSEA-MO) and C_2 =C(nCOSEA-MO, COSEA-MO) for the 3D- and 4D-molecular optimization problem. In general, the C-metric values of the PF of nCOSEA-MO are significantly higher than those of COSEA-MO, thus revealing that more candidate solutions identified by nCOSEA-MO weakly dominate the solutions of COSEA-MO than vice versa within each generation. Figure 1 gives an insight into the number of approximate Pareto optimal solutions identified in the test runs.

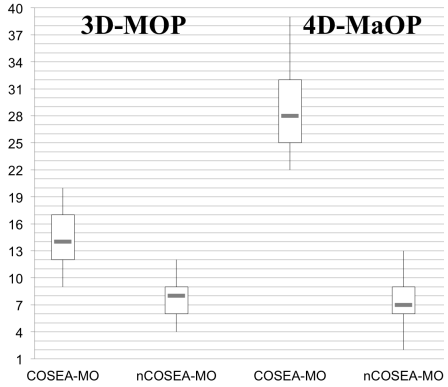

Figure 1: Number of candidate solutions

nCOSEA-MO provides a significantly lower but stable number of candidate solutions, whereas the solutions number of COSEA-MO is generally higher and increases with the problem dimension as a consequence of the Pareto dominance principle being directly applied to the objective values. Consequently, nCOSEA-MO provides a selected number of improved molecules compared to COSEA-MO.

Table 1: C-metric values for 3D-MaOP

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
C_1	0.5	0.3	0.25	0.4	0.25	0.24	0.2	0.17	0.2	0.2
C_2	**0.9**	**1**	**0.93**	**1**	**0.79**	**0.78**	**0.9**	**0.86**	**0.95**	**0.9**

Table 2: C-metric values for 4D-MaOP

	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
C_1	0.4	0.3	0.4	0.4	0.35	0.31	0.4	0.26	0.3	0.25
C_2	**0.7**	**0.5**	**0.6**	**0.7**	**0.57**	**0.64**	**0.7**	**0.79**	**0.84**	**0.82**

4 CONCLUSION

The concept of a selection procedure which applies the Pareto dominance principle on a multi-dimensional indicator-based optimization problem has a high potential to be established in multi- and many-objective optimizations. Future work is intended in direction of high-dimensional molecular optimization problems with up to ten objectives. Furthermore, this concept will be applied in established benchmark problems with indicators referring to the general targets of convergence, diversity and uniformity.

REFERENCES

[1] B. Li, J. Li, and K. Tang. 2015. Many-Objective Evolutionary Algorithms: A Survey. *Comput. Surveys* 48, 1 (September 2015), 13:1 – 13:45. DOI: https://doi.org/10.1145/2792984
[2] S.B. Needleman and C.D. Wunsch. 1970. A General Method Application to the Research for Similarities in the Amino Acid Sequence of Two Proteins. *Journal of Molecular Biology* 48, 3 (1970), 443–453.
[3] S. Rosenthal and M. Borschbach. 2017. Design Perspectives of an Evolutionary Process for Multi-objective Molecular Optimization. *Proc. of the 9th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2017)* LNCS 10173 (March 2017), 529–544.
[4] H. Wang, Y. Jin, and X. Yao. 2016. Diversity Assessment in Many-Objective Optimization. *IEEE Transactions in Cybernetics* PP, 99 (May 2016), 1–13. DOI: https://doi.org/10.1109/TCYB.2016.2550502