Runtime Analysis of Evolutionary Diversity Optimization and the Vertex Cover Problem

[Extended Abstract]

Wanru Gao
Optimisation and Logistics
School of Computer Science
The University of Adelaide
Adelaide, SA 5005, Australia

Mojgan Pourhassan
Optimisation and Logistics
School of Computer Science
The University of Adelaide
Adelaide, SA 5005, Australia

Frank Neumann
Optimisation and Logistics
School of Computer Science
The University of Adelaide
Adelaide, SA 5005, Australia

ABSTRACT

Using evolutionary algorithms to generate a diverse set of solutions where all of them meet a given quality criteria has gained increasing interest in recent years. In order to gain theoretical insights on the working principle of population-based evolutionary algorithms for this kind of diversity optimization a first runtime analysis has been presented by Gao and Neumann [1] on the example problems OneMax and LeadingOnes. We continue this line of research and examine the diversity optimization process of population-based evolutionary algorithms on complete bipartite graphs for the classical vertex cover problem.

Keywords

Combinatorial optimization, Diversity, Theory

1. INTRODUCTION

Evolutionary algorithms (EAs) are widely used for complex problems in various areas such as combinatorial optimization, bioinformatics, and engineering. EAs work with a set of solutions called the population which is evolved during the optimization process. The use of diversity may prevent the algorithms from getting stuck in an optimal solution and set the basis for successful crossover operators. While a population is often used in the context of single-objective optimzation to obtain a single high quality solution, the use of a population has the opportunity to produce a diverse set of multiple solutions which are all of good quality. In this way, a decision maker gets presented several solutions which he can choose from in contrast to just a single one.

We want to study the diversity optimization process of evolutionary algorithms from a theoretical perspective. The first rigorous runtime analysis of maximizing diversity in the decision space has been presented in [1]. In this paper, a variant of the classical $(\mu + 1)$-EA incorporating a mechanism to maximize diversity in the case that solutions meet a given quality threshold has been analyzed. We push the runtime analysis in this context and study special classes of the vertex cover problem. For the vertex cover problem there exist several algorithms that give a 2-approximation of an optimal solution [2, 5]. It is possible to determine the number of existing 2-approximation solutions for the graph under investigation, therefore, we restrict ourselves to classes of graphs that have many solutions that are 2-approximations and present runtime results for the $(\mu + 1)$-EA incorporating diversity maximization for complete bipartite graphs.

GECCO'15, July 11-15, 2015, Madrid, Spain.

2. BACKGROUND

Before discussing the population diversity, we introduce some definitions about diversity in vertex cover problems used in this paper. We consider the vertex cover problem throughout the paper which is given by an undirected graph $G = (V, E)$. The goal is to find a minimum set of nodes $V' \subseteq V$ such that each edge is covered, i.e. for each $e \in E$, $e \cap V' \neq \emptyset$. For our investigations, we assume that the considered algorithms start with a population where each individual is already of desired quality. Our goal is to analyze the runtime until the evolutionary algorithms have obtained a population of good or optimal diversity where all individuals meet the quality criterion.

The solution to the vertex cover problem is represented as binary string, where each 1-bit denotes the existence of corresponding node in a cover set, therefore, we use Hamming distance $H(x, y) = \sum_{i=1}^{n} |x_i - y_i|$, where $x_i, y_i \in \{0, 1\}^n$, to evaluate the difference between two individuals.

According to [3, 4], a diversity measurement should fulfill properties of twinning, monotonicity in varieties and monotonicity in distance. Therefore the diversity of a set of solutions P is defined as follows.

DEFINITION 1. *For a given population P, the population diversity is defined as $D(P) = \sum_{\{x,y\} \in \hat{P} \times \hat{P}} H(x, y)$, where $\hat{P}$ is the set with all distinct solutions in P.*

The contribution of solution x is defined as $c(x) = D(P) - D(P \setminus \{x\})$, where $x \in P$.

The (μ+1)-EA with solution diversity optimization is defined as (μ+1)-EA_D and given in Algorithm 1. In (μ+1)-EA_D, one randomly chosen individual with least contribution of diversity is eliminated from the solution set. The whole process is executed for certain number of generations or until no much improvement in diversity can be made. The initialization process is different for different problems.

Algorithm 1: (μ+1)-$\mathrm{EA_D}$

1 Initialize P with μ n-bit binary strings.
2 Choose $s \in P$ uniformly at random.
3 Produce s' by flipping each bit of s with probability $1/n$ independently from each other.
4 Check whether s' meets the quality criteria or not. If s' fulfils the quality requirement, then add $s'$to P, otherwise go back to step 2.
5 Choose a solution $z \in \{x \in P \mid c(x) = \min_{y \in P} c(y)\}$ uniformly at random. Set $P := P \setminus \{z\}$.
6 Repeat step 2 to 5 until termination criterion is reached.

We study our algorithm in terms of the number of fitness evaluations until it has produced a population with acceptable quality that has the maximal or acceptable diversity $D(P)$. The *expected optimization time* refers to the expected number of fitness evaluations to reach this goal.

3. COMPLETE BIPARTITE GRAPHS

We start by studying complete bipartite graphs. In a complete bipartite graph, the vertices can be split into two sets V_1 and V_2, which are of size ϵn and $(1-\epsilon)n$ respectively. There is an edge between each pair of nodes from set V_1 and V_2. If the nodes in V_1 is indexed from 0 to $\epsilon n - 1$ and nodes in V_2 is indexed from ϵn to $n-1$, the cover set can be represented by a binary string with length n. When $\epsilon < \frac{1}{2}$, a cover set consisting of all the nodes in V_1 is the global optimum of the problem. We use matrix to represent the population where each row represent an individual.

In the vertex cover problem of complete bipartite graph, we focus on the solutions which constitute a 2-approximation of an optimal solution. The population diversity optimization process is conducted on the population after all individuals in the population meet the quality criteria.

The composition of acceptable cover sets depends on the parameter ϵ. Since we focus on the 2-approximation solutions, it is helpful to discuss the relationship between ϵ, $\frac{1}{2}$ and $\frac{1}{3}$.

3.1 $\epsilon < 1/3$

Assume $\epsilon < 1/3$. In order to be 2-approximation to the optimal solution, a cover set should always include every nodes in set V_1 and at most ϵn other nodes in set V_2. The population is initialized with a 2-approximated solution and $(\mu - 1)$ n-bit binary strings randomly chosen from $\{0,1\}^n$.

Taken population diversity into consideration, the (μ+1)-$\mathrm{EA_D}$ aims at finding cover sets of size $2\epsilon n$ and maximize the population diversity. In order to make sure a solution is 2-approximated to the optimal solution set, the leftmost ϵn bits in the bitstring should be set to 1. Then there are at most ϵn bits need to be selected from set V_2, which means in the right $(1-\epsilon)n$ bits, there are at most ϵn 1-bits. The diversity optimization process can be seen as a OneMax problem with population size $\mu = (1-\epsilon)n$ and threshold $v = (1-2\epsilon)n$. The analysis follows the ideas about OneMax of Gao and Neumann [1].

Define P^1_{ini} as a population with μ individuals among which there is at least one individual that is 2-approximation to the optimal cover set.

THEOREM 1. *Let $\frac{1-\epsilon}{\epsilon} \leqslant \mu \leqslant \binom{(1-\epsilon)n}{\epsilon n}$ and $\epsilon < 1/3$, then expected optimization time of (μ+1)-$\mathrm{EA_D}$ on vertex cover problem for complete bipartite graph starting with P^1_{ini} is upper bounded by $O(\mu^3 n^3)$.*

3.2 $\epsilon = 1/3$

If $\epsilon = 1/3$, a 2-approximation cover set can also be composed of all nodes in the larger set. Then there are two types of possible cover sets, typeA and typeB, fulfil the 2-approximation condition, which are all nodes in set V_2 and all nodes in set V_1 together with at most $\frac{1}{3}n$ nodes in set V_2. In order to maximize the population diversity, A should be included in the population, since it contributes the most to the population diversity in the left ϵn columns and with it the right part can still reach optimum diversity. The average number of 0-bits in each column in the left part of matrix is at least $\frac{(\mu-1)(\frac{1}{3}n)}{\frac{2}{3}n} = \frac{\mu-1}{2} < \frac{\mu}{2}$. Then the number of 0-bits and 1-bits in the right $\frac{2}{3}n$ columns should be equal in order to maximize the population diversity. The population with optimal diversity should have solution A and other $(\mu - 1)$ solutions which have equal number of 0-bits and 1-bits in the right $\frac{2}{3}n$ columns which represents the set V_2. The optimum population diversity is $\frac{1}{3}n(\mu-1) + \frac{2}{3}n \cdot \frac{\mu^2}{4}$.

Define P^2_{ini} as a population with μ individuals among which there is one solution has all nodes in set V_2 and at least one individual that includes all nodes in set V_1 and at most ϵn other nodes in set V_2.

THEOREM 2. *Let $\epsilon = 1/3$ and $4 < \mu < \binom{\frac{2}{3}n}{\frac{1}{3}n}$, the expected optimization time of (μ+1)-$\mathrm{EA_D}$ on vertex cover problem for complete bipartite graph starting with P^2_{ini} is upper bounded by $O(\mu^3 n^3)$.*

Acknowledgements

The research leading to these results has received funding from the Australian Research Council (ARC) under grant agreement DP140103400.

References

[1] W. Gao and F. Neumann. Runtime analysis for maximizing population diversity in single-objective optimization. In D. V. Arnold, editor, *Genetic and Evolutionary Computation Conference, GECCO '14, Vancouver, BC, Canada, July 12-16, 2014*, pages 777–784. ACM, 2014.

[2] D. Hochbaum. *Appromixation Algorithms for NP-hard Problems*. PWS Publishing Company, 1997.

[3] T. Ulrich, J. Bader, and L. Thiele. Defining and optimizing indicator-based diversity measures in multiobjective search. In R. Schaefer, C. Cotta, J. Kolodziej, and G. Rudolph, editors, *PPSN (1)*, volume 6238 of *Lecture Notes in Computer Science*, pages 707–717. Springer, 2010.

[4] T. Ulrich, J. Bader, and E. Zitzler. Integrating decision space diversity into hypervolume-based multiobjective search. In M. Pelikan and J. Branke, editors, *GECCO*, pages 455–462. ACM, 2010.

[5] V. Vazirani. *Appromixation Algorithms*. Springer, 2001.