Candidate Oversampling Prefers Two to Tango

Estimation of Distribution Algorithms

David Wallin
University of Limerick
Limerick, Ireland
david.wallin@gmail.com

Conor Ryan
University of Limerick
Limerick, Ireland
conor.ryan@ul.ie

R. Muhammad Atif Azad
University of Limerick
Limerick, Ireland
atif.azad@ul.ie

ABSTRACT

Recent work has enhanced the Evolutionary Bayesian Classifier-based Optimization Algorithm (EBCOA) by oversampling the next generation and identifying promising solutions without actually evaluating their fitness values. In order to model the existing generation, that work considered two *classes* of solutions, that is, high performing solutions (H-Group) and poorly performing solutions (L-Group). In this study, we test the utility of using two classes instead of using a single class, as is the norm in standard Estimation of Distribution Algorithms (EDAs). Our results show that a dual class model is preferable when oversampling is used.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search; G.1.6 [**Numerical Analysis**]: Optimization

General Terms

Algorithms

Keywords

EDA, Estimation of Distribution, Evolutionary Computation, Probabilistic Model, Probabilistic Model-Building

1. INTRODUCTION

Estimation of Distribution Algorithms (EDAs) [1] combines an evolutionary search with a selection and sampling mechanism based on probabilistic modeling. EDAs have shown themselves capable of solving hard problems such as deceptive and hierarchical problems.

To drive the search forward, EDAs model on a subset of promising solutions. Typically, the information is gathered only from the good solutions in a population. However, a growing number of algorithms, such as EBCOA [2], try to, in addition, extract useful information from the poor-performing solutions. This is done by transforming the task of optimisation into that of classification between different classes of candidate solutions, e.g., top performing candidates and low-performing ones. The model tries to capture the differentiating characteristics of solutions that belong to different classes, and subsequently use this information when instantiating. This is the underlying motivation behind the approach taken by EBCOA.

A recently introduced addition [4] to the EBCOA algorithm, *candidate oversampling*, identifies promising solutions from an oversized population. The identification is performed without evaluating their actual fitness values.

In this paper we will evaluate the method proposed in [4] and contrast its performance on the EBCOA where dual classes is modeled, to the performance on an EDA where single class model is used.

We will show that the performance of EBCOA and that of an EDA using an equivalent, but single class model, is comparable. However, when candidate oversampling is added, the use of the dual class model found in EBCOA becomes essential to solve some of the problems.

1.1 Estimating sample suitability

The method introduced in [4] proposed the use of oversampling from a model. A metric gives an estimation of each sample's suitability, and then truncates the collection of samples back to the original size of the population, keeping those that are considered promising.

In our study, we will use the *Weaken Likelihood* metric from [4], which consistently, and surprisingly, outperformed the other metrics, as well as EBCOA with sampling fixed to a single class (namely the H-group).

Weaken Likelihood (WL) is based on that an instantiation's *likelihood* of belonging to a class c can be measured as: $\theta_t(\mathbf{x}, c) = \prod_{i=1}^{n} \sqrt{p_t(x_i|\pi_{x_i}, c)}$, where x_i is an instantiation of the i^{th} variable, given its set of parents π_{x_i}, and the class c. The square root guards against numerical underflows that can arise when multiplying the many probabilities.

When the WL metric is applied to the data set D''_{t+1}, the samples classified as least likely to be in the H-group are selected: $D'_{t+1} \leftarrow min_R(\theta_t(\mathbf{x}_1, H), \ldots, \theta_t(\mathbf{x}_{kR}, H))$, where min_R is a function that selects and returns R samples with the smallest input values. The resulting data set D'_{t+1} contains the candidate solutions that will be fitness evaluated. The use of candidate oversampling will thus perform the same number of fitness evaluations as would be the case when sampling the population D'_{t+1} directly.

2. EXPERIMENTS

The role of the model in EBCOA is to both model partitions of the population as well as generating new samples. The standard EBCOA [2] used dual classes for both. We can refer to this as a C2C2 method. Coequally, the work on candidate oversampling [4] used a single class sampling, but

GECCO'11, July 12–16, 2011, Dublin, Ireland.
ACM 978-1-4503-0690-4/11/07.

maintained two classes for modeling: C1C2. In this work we compare the C1C2 approach to a more EDA-like C1C1 method, which will in essence work like an EDA based on a bivariate dependency model.

We will test the performance of WL for cases both with and without the introduction of diversity into the sampling phase. For this purpose we employ the *Sampling-Mutation* (SM) [3] operator in some of our experiments with a probability of 0.01, and are marked "m01". Similarly, the "wl" tag indicates the use of Weaken Likelihood.

3. RESULTS

On a 256-bit HIFF problem with an optimal fitness value of 2304, we use a population size of 1600 and let it run for 200 generations. The results of the experiments are shown in Figure 1, where the optimum fitness is marked with a dotted line. In the case of C1C1, it is apparent that none of the results indicate that the problem is solved. The results are summarised in Table 1 along with the corresponding standard deviation values (s).

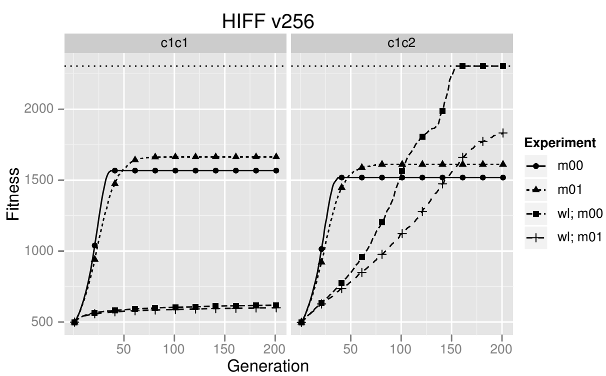

Figure 1: Fitness for the experiments on the 256-bit HIFF problem.

Settings	C1C1	s	C1C2	s
m00	1567.46	±111.02	1518.40	±103.91
m01	1663.60	±126.57	1611.20	±115.07
m00-wl	618.40	±10.39	2304.00	0.00
m01-wl	600.12	±8.88	1832.70	±83.04

Table 1: Comparison of the results on the 256-bit HIFF problem between the two approaches.

The experiments, m00 and m01, that use the EDA-like setup without WL, both show the typical signs of a fast convergence with a rapid loss in population diversity. The result is a fitness fixation from which there is no escape. The WL experiments, using C1C1, show very little improvement from the initial fitness value. This is in stark contrast to the same experiments using a C1C2 algorithm.

The results for another hierarchical problem HXOR is similar to those on HIFF, and are summarised in Table 2. Table 3 show the results on a 200-bit Ising problem.

The abysmal performance of WL in a C1C1 setting does not extend to all problems. Experiments on other problem domains showed that the performance between C1C1 and C1C2 were similar, both with and without the use of candidate oversampling.

Settings	C1C1	s	C1C2	s
m00	1687.28	±139.11	1643.70	±124.73
m01	1775.36	±139.78	1719.40	±126.98
m00-wl	602.52	±9.58	2304.00	0.00
m01-wl	586.80	±9.24	1817.60	±52.80

Table 2: Comparison of the results on the 256-bit HXOR problem between the two approaches.

Settings	C1C1	s	C1C2	s
m00	177.16	±5.13	176.88	±4.45
m01	182.32	±4.59	182.16	±3.98
m00-wl	156.88	±2.10	200.00	0.00
m01-wl	111.80	±1.19	200.00	0.00

Table 3: Comparison of the results on the 200-bit Ising problem between C1C1 & C1C2.

4. CONCLUSIONS AND FUTURE WORK

In this paper we set out to test candidate oversampling using the Weaken Likelihood metric, as proposed in [4], using a probabilistic model where no class-information is used neither in the sampling phase, nor in the building of the model.

From the experiments that compared C1C1 with C1C2 with no candidate oversampling, we conclude that there is very little difference between the two. However, when candidate oversampling with WL is added, the picture is quite different. The results show that WL works best when used in conjunction with C1C2.

The question why WL works so well in a C1C2 setting, and work so poorly with C1C1 is worth further investigation.

As it stands, C1C2 seems to be a more stable foundation for candidate oversampling and WL.

5. REFERENCES

[1] Pedro Larrañaga and José A. Lozano, editors. *Estimation of Distribution Algorithms - A New Tool for Evolutionary Computation*, volume 2 of *Genetic Algorithms and Evolutionary Computation*. Springer-Verlag, 2001.

[2] Teresa Miquélez, Endika Bengoetxea, and Pedro Larrañaga. Evolutionary Computation based on Bayesian Classifiers. *International Journal of Applied Mathematics and Computer Science*, 14(3):335–349, 2004.

[3] David Wallin and Conor Ryan. On the Diversity of Diversity. In *Proceedings of the 2007 Congress on Evolutionary Computation (CEC2007)*, pages 95–102. IEEE, 2007.

[4] David Wallin and Conor Ryan. Using Over-Sampling in a Bayesian Classifier EDA to Solve Deceptive and Hierarchical Problems. In Andy Tyrrell, editor, *2009 IEEE Congress on Evolutionary Computation*, Trondheim, Norway, 18-21 May 2009. IEEE Computational Intelligence Society, IEEE Press.