A Clustering Entropy-driven Approach for Exploring and Exploiting Noisy Functions

Shih-Hsi Liu
Department of CIS
Univ. of Alabama at Birmingham
Birmingham, AL 35294, USA
liush@cis.uab.edu

Marjan Mernik
Faculty of EECS
Univ. of Maribor
2000 Maribor, Slovenia
marjan.mernik@uni-mb.si

Barrett R. Bryant
Department of CIS
Univ. of Alabama at Birmingham
Birmingham, AL 35294, USA
bryant@cis.uab.edu

ABSTRACT

Linear, Gaussian, fitness proportional, clustering, and Rosca entropies are succinct measures of diversity that have been applied to balance exploration and exploitation in evolutionary algorithms. In previous studies, an entropy-driven approach using linear entropy explicitly balances and/or searches optimal solutions for the selected unimodal and multimodal functions excluding noisy functions. This paper investigates the reasons for such an exception and introduces a clustering entropy-driven approach to solve the problem. Such an approach provides a coarse-grained diversity measure that filters the noise of functions, varies cluster size and categorizes individuals at the genotype level. The experimental results show that the clustering entropy-driven approach further improves the searching results of noisy functions by one more degree.

Keywords
Cluster, Entropy, Exploration, Exploitation

1. INTRODUCTION

An Evolutionary Algorithm (EA) is a population-based metaheuristic and stochastic optimization process comprising two major aspects for a search space: exploration and exploitation [2]. To acquire uncovered promising solutions, exploration discovers potential offspring by directing the search space to entirely new regions; and to retain visited promising solutions, exploitation determines potential regions of the search space to be explored next by utilizing such visited information. How good (i.e., optimal solutions) and how fast (i.e., convergence rate) an evolutionary process performs is, hence, concluded by the correlation of exploration and exploitation. Despite that, how and when to control either exploration or exploitation to dominate the evolutionary process and then to maximize the virtue of the correlation is still ongoing research.

In [14, 15], we concentrated on controlling and balancing exploration and exploitation during the evolutionary process using a domain-specific language, PPC/EA [13]. This approach used five types of entropy as succinct diversity measures and mutation rate (\(p_m\)) as an exploration measure to determine the trends of balance during the evolutionary process: if an entropy value is higher than a pre-defined boundary value, more exploitation is performed by decreasing \(p_m\); if an entropy value is lower than the boundary value, more exploration is facilitated by increasing \(p_m\). Additionally, the five entropies (i.e., linear, Gaussian, fitness proportional, clustering [15], and Rosca [17]) utilized different classification techniques to categorize individuals into different subgroups (i.e., fitness classes) and computed the entropy values based on the probability of individuals occupied in each subgroup [17]. For unimodal and multimodal functions selected from [20], the experimental results in [15] showed that the entropy-driven approach using linear entropy was outstanding in explicitly balancing and/or searching optimal solutions except noisy quartic functions, i.e., functions with a quartic formula with a random number generator. The linear entropy-driven approach also outperformed the Fogarty [4], Schaffer [10], and 1/5 success rule approaches in terms of convergence rate for unimodal functions and in terms of mean best fitness results for multimodal functions. The main reason for the fitness outperformance is that the linear entropy-driven approach facilitated exploration to remain active at the later phase of the evolutionary process and resulted in discovering better solutions late in the process.

Nevertheless, as described above, the linear entropy-driven approach for noisy quartic functions did not generate distinguished results. The experimental figures in [15] (e.g., Figure 3 in this paper) showed that no entropy revealed apparent diversity information that assisted the entropy-driven approach in balancing exploration and exploitation explicitly. There are two major reasons for such a phenomenon applied to noisy quartic functions. First, the classification methods of linear, Gaussian, fitness proportional, and Rosca entropies are at the phenotype level and too fine-grained such that very subtle fitness differences among individuals incurred high and steady entropy values. For example, Rosca used fitness values to classify individuals [17]. The random number generator in \(f_r\) forced each individual to possess a unique fitness value, and thus the Rosca entropy value was always 2 during the evolutionary process. Second, despite having a coarse-grained classification method at the
genotype level to filter out noise, clustering entropy did not offer a suitable diversity measure due to an inappropriate fixed size of subgroups (i.e., clusters) in [15].

Based on the experimental results of the linear entropy-driven approach and the observations of the remaining four entropies for noisy quartic functions, this paper introduces a clustering entropy-driven approach. Such an approach utilizes clustering entropy to balance the evolutionary process and configures different cluster sizes and the boundary values of exploration and exploitation. The experimental results show that the clustering entropy-driven approach further improves the searching results of noisy quartic functions in one more degree compared with the linear entropy-driven approach.

The paper is organized as follows. The next section summarizes the related work. Section 3 presents the clustering entropy-driven approach for noisy quartic functions. Section 4 shows the experimental results. Finally, the last section concludes the paper and states the future work.

2. RELATED WORK

Optimal balance between exploration and exploitation has been mainly controlled by determining the diversity of a population during an evolutionary process. Three major aspects of this topic have been categorized: diversity persistence, diversity measurements, and diversity learning [15]. Many techniques have been introduced to tackle this topic. For space consideration, only a few techniques are presented.

For diversity persistence, the main purpose of such an aspect is to maintain the diversity of a population at different stages of an evolutionary algorithm. For example, the selection process can be used to control the level of exploration or exploitation by varying the selection pressure [2]. More exploitation is applied if the selection pressure is high, and more exploration is performed if the selection pressure is low. In [9], a modified evolutionary strategy with a diversity-based selection pooling scheme is introduced. Such a scheme modulates the number of parents entering the selection pool based on several diversity measures.

For diversity measurements, standard deviation and Euclidean distance are commonly used in EAs. Rosca [17] investigated entropy, a representation of diversity, in EAs. His experiments showed that populations appeared to be stuck in local optima when entropy did not change or decrease monotonically in successive generations. Rosca used fitness values in a population to define entropy and free energy measures. In [19], a distance-to-average-point measure was introduced to alternate exploration and exploitation in the diversity-guided evolutionary algorithm.

For diversity learning, the Genetic Algorithm using Self-Organizing Maps (GOSAM) [1] utilizes neural network concepts to achieve such a prospect. Initially, a training set of a Self-Organizing Map (SOM) [18] is exercised as a reference of diversifying individuals. Two tables that store current population diversity and search history are analyzed and visualized using the SOM. Users can easily discover the regions of search space to be explored. Balance between exploration and exploitation is controlled by a balance function which indicates the number of neurons that should be activated at the current generation. A reseeding operator that preserves the exploration power is driven by the difference between expected and current activations of neurons.

For all the techniques described above, none of them addresses how to present diversity measures properly and how to learn diversity during the evolutionary process in a noisy environment. This paper concentrates on these two hows to solve the problems mentioned in Section 1. Please note that the paper does not focus on the diversity persistence aspect, because our goal is to obtain and utilize real diversity information from a noisy environment.

Additionally, there have been a variety of studies for EAs in noisy environments using non-diversity approaches [11]. For example, Fitzpatrick and Greffenstette [3] duplicated individuals and then averaged their fitness values to reduce noise interference. Goldberg, Deb, and Clark experimented with a noisy function in [6] and later introduced a control map [5] for parameter control. However, both approaches concentrated on the quality of convergence instead of mean best fitness results as presented in this paper.

3. CLUSTERING ENTROPY-DRIVEN APPROACH

For space consideration, only a few techniques are presented. Many techniques have been introduced to tackle this topic. For example, Fitzpatrick and Greffenstette [3] duplicated individuals and then averaged their fitness values to reduce noise interference. Goldberg, Deb, and Clark experimented with a noisy function in [6] and later introduced a control map [5] for parameter control. However, both approaches concentrated on the quality of convergence instead of mean best fitness results as presented in this paper.

Before delving into the problems, the linear entropy-driven approach and two of its experimental results are revisited. As described before, the linear entropy-driven approach is superior in the selected unimodal and multimodal functions.
except noisy quartic function. Figures 1 and 2 show the experimental results of the Sphere Model ($f_1$), a unimodal function, and Ackley’s function ($f_{10}$), a multimodal function with many local minima, respectively. The explicit balance between exploration and exploitation of both figures can be observed using curves $B$ (best fitness), $A$ (average fitness), $W$ (worst fitness), $S$ (standard deviation), $D$ (diversity), $E$ (linear entropy), $G$ (Gaussian entropy), $P$ (fitness proportional entropy), $R$ (Rosca entropy), $C$ (clustering entropy), and $p_m$ (mutation rate).

In Figure 1, curves $E$, $P$, and $R$ steeply decline at the early phase. When curve $E$’s value is between its midpoint and upper bound, $p_m$ is decreased to balance exploitation against exploration. As curve $E$’s value is between its lower bound and midpoint, exploration outperforms exploitation by rising $p_m$. The characteristics of balance between exploration and exploitation in Figure 2 are similar to Figure 1. Because of the later exploration on the search space of local minima, fitness values are still slightly improved at the ending phase.

However, in a noisy environment, different parameters will be influenced by noise in various magnitudes at a specific level (i.e., genotype or phenotype). This section uses noisy quartic functions to discuss such influence. Based on the discussion, the clustering entropy-driven approach is introduced to solve the problems occurring in noisy quartic functions. The following equation is the noisy quartic function ($f_7$) from [20].

$$f_7(x) = \sum_{i=1}^{d} i \ast x_i^4 + \text{random}[0, 1]$$  \hspace{1cm} (1)

where $x_i \in [-1.28, 1.28]$, $d$ (dimension) = 30, and $\text{min}(f_7) = f_7(0,...,0) = 0$.

For fitness parameters (i.e., $B$, $A$, and $W$), noise generated by the random number generator shifts the fitness values up to 1 (see eq. 1). Such additions affect the process of searching optimal solutions. The solutions interfered by noise, however, can be diluted by averaging the parameters from sufficient numbers of experiments. For diversity parameters (i.e., $S$ and $D$), standard deviation computed from fitness values possesses the same characteristic as above. Diversity is the Euclidean distance among all individuals. The random number generator does not have any noise influence on this parameter, because diversity and noise are computed at the genotype and phenotype levels, respectively. For the entropy parameters (i.e., $E$, $G$, $P$, $R$, and $C$), however, noise influences these parameters in different magnitudes pertaining to how each entropy classifies a population and how many fitness classes (i.e., subgroups) are partitioned: $E$ classifies individuals based on evenly partitioned intervals between best and worst fitness values with a predefined fitness class size; $G$ utilizes Gaussian distribution for category; $P$ groups individuals by the proportions of fitness values; $R$ classifies individuals using fitness values; and $C$ will be introduced later.

Figure 3 shows the experimental results using the linear entropy-driven approach. In the figure, the fitness class sizes of linear entropy ($n$) and clustering entropy ($k$) are 100 and 5, respectively. The boundary value for linear entropy ($EB$) is 0.5. As shown in Figure 3, curves $R$ and $P$ remain at 2 whereas curves $E$ and $G$ offer very limited diversity information for noisy quartic function. Curve $C$ decreases gradually, similar to curve $S$ and $D$, and crosses the boundary value at generation 400. Unfortunately, because the experiment is a linear entropy-driven approach, the inactive curve $E$, whose value is always above the boundary value, can never facilitate exploration during the entire evolutionary process. The advantages of the entropy-driven approach are, hence, never taken.

Clustering entropy adopts a self-adaptive classification technique, which partitions data based on specific similarity criteria. Such a technique utilizes the K-Means clustering algorithm [7] to group individuals into a predefined fixed $k$ fitness classes, where $1 \leq k \leq \text{the population size (Popsize)}$. $k$ individuals are initially chosen as the centroids of clusters. The iterative classification process computes the Euclidean distances between individuals and centroids and then assigns individuals to the centroid having the shortest Euclidean distance. After each intermediate classification step, the centroids are updated by averaging the genotypes of all the individuals in the same fitness class (i.e., cluster or subgroup). The iterative classification process will be stopped as the similarity criterion is satisfied, i.e., the centroids are not updated anymore in the experiments in Section 4.

The key characteristic which yields clustering entropy applicable to the entropy-driven approach in a noisy environment is its iterative classification process. As mentioned above, such a process employs Euclidean distance to categorize individuals. Similar to parameter $D$, Euclidean distance computed at the genotype level will not include noise added at the phenotype level. The classification process, therefore, filters out the interference driven by noise and results in a coarse-grained diversity measure for clustering entropy.

Figure 4 shows the clustering entropy-driven approach written in PPCeA. The initial values of necessary parameters are defined in lines 1 and 3: Epoch is the stride to adjust parameters during the evolutionary process; Round is the total number of experiments for the same function; Maxgen is the maximum generation number of noisy quartic function; $p_m$ and $p_c$ are mutation and crossover rates; and clentropy is clustering entropy designed in PPCeA. Lines 4 and 6 respectively initialize the population and perform the evolutionary process of genetic algorithms. From lines 7 to 9, more exploitation will be carried out by decreasing $p_m$, if the clustering entropy value is larger than the predefined boundary value. Conversely, from lines 10 to 12, more explo-
1. EXPERIMENTAL RESULTS

This section presents the experimental results. Figures 5 and 6 exercise the clustering entropy-driven approach with different boundary values.

![Figure 5: The clustering entropy-driven approach with n = 5, k = 10, and boundary value (CB) = 0.5.](image)

In Figure 5, 10 centroids (k = 10) are selected for classification. Such a process presents an explicit balance between exploration and exploitation. Before generation 850, as the clustering entropy value is higher than the boundary value, the evolutionary process tends to exploit visited regions to obtain optimal solutions. From generations 850 to 2100, exploitation and exploration are equally dominant between each other. Curves B, A, W, S, and D slightly decline during this phase. Lastly, exploration is awakened by the low clustering entropy value to avoid being stuck at local optima and to discover promising even better solutions.

![Figure 6: The clustering entropy-driven approach with n = 5, k = 10 and boundary value (CB) = 0.7.](image)

Figure 6 tackles the influence of boundary values for balancing exploration and exploitation. The higher boundary value (0.7) carried out is a revised version of Figure 5. pm is increased 1200 generations earlier than the one in Figure 5 due to the higher boundary value. The exploration process, therefore, has sufficient time to explore a better solution.

<table>
<thead>
<tr>
<th>Entropy</th>
<th>(n,k)</th>
<th>EB/CB</th>
<th>B</th>
<th>S</th>
<th>CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>(100,5)</td>
<td>0.5</td>
<td>3.06E-2</td>
<td>1.55E-4</td>
<td>1771</td>
</tr>
<tr>
<td>Linear</td>
<td>(50,5)</td>
<td>0.5</td>
<td>3.69E-2</td>
<td>1.46E-4</td>
<td>2087</td>
</tr>
<tr>
<td>Linear</td>
<td>(10,5)</td>
<td>0.5</td>
<td>3.11E-2</td>
<td>1.49E-4</td>
<td>1862</td>
</tr>
<tr>
<td>Linear</td>
<td>(5,5)</td>
<td>0.5</td>
<td>2.72E-2</td>
<td>1.56E-4</td>
<td>2405</td>
</tr>
<tr>
<td>Clustering</td>
<td>(5,20)</td>
<td>0.5</td>
<td>2.15E-2</td>
<td>4.78E-4</td>
<td>2888</td>
</tr>
<tr>
<td>Clustering</td>
<td>(5,15)</td>
<td>0.5</td>
<td>1.98E-2</td>
<td>5.36E-4</td>
<td>2912</td>
</tr>
<tr>
<td>Clustering</td>
<td>(5,10)</td>
<td>0.5</td>
<td>2.02E-2</td>
<td>4.76E-4</td>
<td>2994</td>
</tr>
<tr>
<td>Clustering</td>
<td>(5,8)</td>
<td>0.5</td>
<td>1.71E-2</td>
<td>5.51E-4</td>
<td>2846</td>
</tr>
<tr>
<td>Clustering</td>
<td>(5,6)</td>
<td>0.5</td>
<td>1.16E-2</td>
<td>5.92E-4</td>
<td>2943</td>
</tr>
<tr>
<td>Clustering</td>
<td>(5,5)</td>
<td>0.5</td>
<td>8.84E-3</td>
<td>7.35E-4</td>
<td>2937</td>
</tr>
<tr>
<td>Clustering</td>
<td>(5,4)</td>
<td>0.5</td>
<td>1.06E-2</td>
<td>1.63E-4</td>
<td>2818</td>
</tr>
<tr>
<td>Clustering</td>
<td>(5,3)</td>
<td>0.5</td>
<td>1.74E-2</td>
<td>5.82E-4</td>
<td>2943</td>
</tr>
<tr>
<td>Clustering</td>
<td>(5,10)</td>
<td>0.7</td>
<td>9.90E-3</td>
<td>7.89E-4</td>
<td>2914</td>
</tr>
<tr>
<td>Clustering</td>
<td>(5,10)</td>
<td>0.25</td>
<td>3.03E-2</td>
<td>2.56E-4</td>
<td>2884</td>
</tr>
<tr>
<td>Clustering</td>
<td>(5,5)</td>
<td>0.25</td>
<td>2.67E-2</td>
<td>3.64E-4</td>
<td>2969</td>
</tr>
</tbody>
</table>

To summarize, the nearly flat curve E in all linear entropy-driven experiments using different settings is not an appropriate diversity measure and control mechanism to perform the balance between exploration and exploitation for noisy quartic function. Even though the decrease of the fitness class size for linear entropy improves the results slightly, balancing directions may be too chaotic in different experi-
ments due to the nearly flat curve (shown in Table 1\(^1\)). For the clustering entropy-driven experiments, a sharp curve \(C\) solves the problem of non-deterministic balance directions occurring in the linear entropy-driven approach and helps to discover better solutions compared with the linear entropy-driven approach, as shown in Figures 5 and 6. Despite having a sharp curve \(C\), a clustering entropy-driven experiment should also have a suitable boundary value to facilitate explicit balance between exploration and exploitation. Inappropriate boundary value settings will affect the right time to switch between exploration and exploitation. Additional statistical results of other parameter settings are also presented in the table (figures available in [12]). The table shows that the best parameter settings out of the listed 15 experiments are \((k = 5, CB = 0.5)\) and \((k = 10, CB = 0.7)\) using the clustering entropy-driven approach. Finally, although the clustering-driven approach results in worse convergence generation values compared with the linear-entropy approach, these values are still better or same to those of Schaffer, Fogarty, and 1/5 success rule approaches in [15].

5. CONCLUSIONS

In a noisy environment, different parameters will be influenced by noise in various magnitudes at the genotype or phenotype level. For the entropy parameters, the magnitudes influenced by noise pertain to how each entropy classifies a population and how many fitness classes are categorized. The entropy-driven approach [14, 15] utilizes such diversity measures to balance exploration and exploitation of an evolutionary process. However, due to the interference of noise, the entropy-driven approach using linear entropy does not perform well for noisy quartic functions. This paper uses clustering entropy, as a coarse-grained diversity measure, with a self-adaptive classification technique to overcome the fine-grained classification problems in [15]. The paper also exercises the clustering entropy-driven approach using different sizes of fitness classes and boundary values. The experimental results show that fitness class sizes decide how sharp a clustering entropy curve is and boundary values determine when exploration and exploitation is alternated. The correlation of the two parameters should be properly set up in order to obtain optimal solutions of noisy quartic functions. Our future direction is to introduce a deterministic or an adaptive approach to adjust fitness class sizes and/or boundary values on-line based on the current entropy and fitness values for different benchmark functions. Experimenting on noisy functions using evolutionary strategies (e.g., CMA-ES [8]) and using genetic algorithms with non-diversity parameters (e.g., control map [5]) are also planned.

6. REFERENCES