Novel Virtual Fitness Evaluation Framework for Fitness Landscape Learning Evolutionary Computation

Taku Hasegawa, Kaname Matsumura, Kaiki Tsuchie,
Naoki Mori, Keinosuke Matsumoto
Osaka Prefecture University
1-1 Gakuencho, Sakai city Osaka, 599-8531, JAPAN
hasegawa@ss.cs.osakafu-u.ac.jp

ABSTRACT

Introducing the machine learning technique into evolutionary computation (EC) is one of the most important issues to expand EC design. In this paper, we proposed a novel method that combines the genetic algorithm and support vector machine to achieve the imaginary evolution without real fitness evaluations.

Categories and Subject Descriptors

I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search—Heuristic methods

General Terms

Algorithms

Keywords

Genetic algorithms; Fitness evaluation; Surrogate model/fitness approximation; Selection; Machine learning

1. INTRODUCTION

The evolutionary computations (ECs) are search and optimization algorithms based on the mechanism of natural evolution. The cost of fitness evaluations in EC becomes a critical issue in several problems, so reducing the number of fitness evaluations is a very important issue. To solve this problem, there have been reported several researches about evaluating fitness approximately[1] and authors have researched imaginary fitness evaluation in EC by machine learning[2]. However, lots of methods utilize the fitness evaluation by machine learning as supplementary methods.

In this study, we extend our proposed method[2] to the novel evolutionary computation framework called the " Fitness Landscape Learning Evolutionary Computation (FLLEC)", which can learn the fitness landscape by machine learning and estimate fitness approximately in order to evolve EC population imaginary. Although we adopted artificial neural network to learn the fitness landscape in our old research[2], we combined the support vector machine (SVM) and GA as typical example of FLLEC in this paper. The research combining the EC and SVM has been proposed[3], however this study only focused on evolutionary strategy (ES) which utilized the real number gene and avoided the evolution by fitness evaluation based on machine learning. On the other hand, the population has evolved based on imaginary fitness evaluation positively in FLLEC. To show the effectiveness of our methods, the computational experiments are carried out taking several combinational problems as examples.

GECCO'14, July 12–16, 2014, Vancouver, BC, Canada.
ACM ACM 978-1-4503-2881-4/14/07.
http://dx.doi.org/10.1145/2598394.2598496.

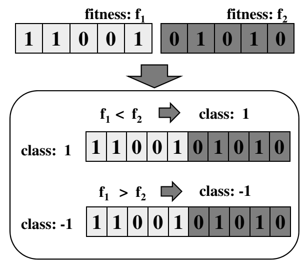

Figure 1: Outline of SVM

2. AIR GA

The FLLEC with GAs and SVM is called "air GA". In this research, the input of the SVM is a two individuals genotype based information and the output is about class labels, superiority or inferiority. Figure 1 shows the outline of our SVM.

The trained SVM is utilized in selection operator in air GA. One of the most important features of the proposed method is that the population is evolved by fitness evaluation of machine learning.

There are two different types of air GA. The first one switches the selection based on direct fitness evaluation and SVM evaluation by interval generations. The second one executes the selection based on both direct fitness evaluation and SVM evaluation in the same generation. In this paper, we use the second one as air GA

The algorithm of this method is following:

1. Set the number of regular evaluations M, and the interval of relearning T. If population size is N, $M \leq N$. If $M = N$, this algorithm is equal to the normal GA.

2. Execute normal GA in the first generations.
3. Train the SVM with individuals and fitness data obtained in step 2.
4. Estimate the real fitness of M new genotypes and execute GA with tournament selection by normal evaluation and air evaluation as mentioned above.
5. Before selection of each T generation $(1, 1+T, 1+2T, \cdots)$, SVM is relearned by new individuals data obtained in step 4 and new SVM is utilized in selection on current generation. The maximum number of new individuals obtained previous interval is MT.
6. Finish the algorithm if the number of fitness evaluations becomes the maximum value.

3. EXPANSION METHOD OF LEARNING IN AIR GA

In air GA, SVM is always updated after relearning SVM. To recognize current genotype space correctly, SVM has to be relearned by individuals obtained in recent populations. However, there exist lots of old SVMs in air GA search process which are only used once. We did not reuse old SVMs, but those are also important to avoid over fitting. Therefore we expand the way of learning in air GA to consider not only latest SVM but old SVMs by introducing ensemble learning.

3.1 Ensemble Learning

Ensemble Learning is learning algorithms that construct a set of classifiers and then classify unknown data points by taking a weighted vote of each classifier's predictions. To apply ensemble learning to air GA, two type of ensemble learning have to be considered. In air GA, we regard that training data set is created by sampling with replacement. This feature is very similar to bagging type ensemble learning. However, because there exists the bias of selection pressure, we cannot consider all SVMs as equal. The main purpose of air GA is not creating perfect SVM but finding the optimal solution. This indicates current SVM is more important than old ones. To reflect this feature, we introduce weight parameter to SVMs like boosting which is one of the popular ensemble methods. We have to notice that this weight is not depending on accuracy rate but on the mean fitness of training data.

3.2 Damping factor

In selection of the expansion method, all SVMs obtained in past generations are utilized. The input of SVM is a string composed of the genotypes of two individuals x_1 and x_2. We represent the i-th SVM's output for classifying the fitness order of x_1 and x_2 as $g_i(x_1, x_2)$. Proposed classifier $g(x_1, x_2)$ is represented by the following equation:

$$g(x_1, x_2) = \sum_{i=0}^{T} \alpha^{T-i} g_i(x_1, x_2) \qquad (1)$$

$$\text{where} \begin{cases} 0.5 \leq \alpha \leq 1.0 \\ \alpha = 0 \end{cases}$$

The air GA only utilizing the latest SVM is equal to the case where $\alpha = 0$ in Eq. (1).

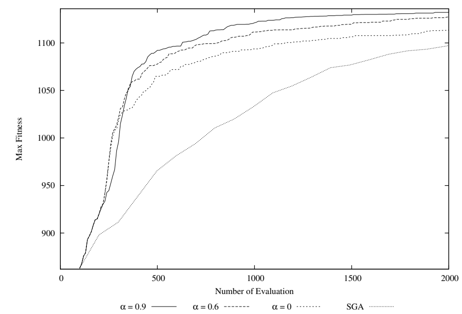

Figure 2: Max fitness with fitness evaluations (Knapsack problem)

4. EXPERIMENTS

In order to show the effectiveness of air GA with ensemble learning(ensemble air GA), we compare the performance of normal air GA, ensemble air GA and SGA after parameter tuning taking knapsack problem as an example.

Figure 2 shows the results. In the figure, normal air GA is represented as the case where $\alpha = 0$. In Figure 2, the abscissa shows the number of fitness evaluations and the ordinate shows the average of max fitness in all the trials.

Figure 2 shows that max fitness is improved remarkably in proportion to α in each generation. The reason is that lots of poor individuals are created in knapsack problem because of our over limit weight setting, and ensemble learning is effective to classify such kinds of individuals.

5. CONCLUSION

In this paper, we proposed the novel EC framework called FLLEC by extending our proposed method[2]. We also showed that the number of fitness evaluations was largely reduced by FLLEC compared to that of SGA.

The way of setting restructuring interval and analyzing the search performance in various parameters and various learning algorithm are important further studies.

Acknowledgment

This work was supported by JSPS KAKENHI Grant, Grant-in-Aid for Scientific Research(C), 2500208, 2010-2013

6. REFERENCES

[1] Y. Jin : A comprehensive survey of fitness approximation in evolutionary computation, Soft Computing, Springer, vol. 9, no. 1, pp 3-12 (2005)

[2] N. Mori, T. Kude, K. Matsumoto : On Genetic and Environment Identifying Approaches to Adaptive to Changing Environments, Workshop program of GECCO'2001,pp. 39-42 (2001)

[3] I. Loshchilov, M. Schoenauer, M. Sebag : Intensive Surrogate Model Exploitation in Self-adaptive Surrogate-assisted CMA-ES (saACM-ES), Proceedings of the 2013 ACM/SIGEVO GECCO-2013, pp. 439-446, (2013).