Importance of Finding a Good Basis in Binary Representation

Junghwan Lee
Dept. Comp. Sci., Kwangwoon Univ.
Seoul, Republic of Korea
jazz4rabbit@gmail.com

Yong-Hyuk Kim
Dept. Comp. Sci., Kwangwoon Univ.
Seoul, Republic of Korea
yhdfly@kw.ac.kr

ABSTRACT

In genetic algorithms, the importance of the basis for representation has been well known. In this paper, we studied the effect of a *good basis* in binary representation, and resultantly we could show that a good basis improves the performance of search algorithms. A complicated problem space may be transformed into a linearly-separable one via a change of basis. We had experiments on search performance. Finding a good basis from all the bases may not be practical, because it takes $O(2^{n^2})$ time, where n is the length of a chromosome. However, we used a genetic algorithm to find a good basis, to correctly investigate how a basis affects the problem space. We also conducted experiments on the NK-landscape model as a representative computationally hard problem. Experimental results showed that changing basis by the presented genetic algorithm always leads better search performance on the NK-landscape model.

CCS CONCEPTS

• **Theory of computation** → *Evolutionary algorithms*; • **Computing methodologies** → Genetic algorithms;

KEYWORDS

basis change, binary representation, meta-genetic algorithm, general linear group, NK-landscape

ACM Reference Format:

Junghwan Lee and Yong-Hyuk Kim. 2018. Importance of Finding a Good Basis in Binary Representation. In *GECCO '18 Companion: Genetic and Evolutionary Computation Conference Companion, July 15–19, 2018, Kyoto, Japan.* ACM, New York, NY, USA, Article 4, 2 pages. https://doi.org/10.1145/3205651.3208774

1 INTRODUCTION

A matrix A is called *binary* if $A \in M_{n \times n}(\mathbb{Z}_2)$. Binary matrices can be used to deal with the adjacency of a graph. They can also be used to represent a change of basis of a vector space over $\mathbb{Z}_2$. Gene reordering can be considered as a special case of a change of basis [3] and helps improve genetic algorithm performance [5, 7].

In this paper, we conduct an empirical study on the influence of a basis in a genetic algorithm with binary encoding. The basis can help genetic algorithms create and preserve high-quality schemata. As a result, it improves the performance of genetic algorithms. We call such a basis a *good basis*. However, since it takes $O(2^{n^2})$ time to search all bases, we used a meta genetic algorithm to find a good basis. We conducted experiments on the NK-landscape model. We tried to solve an NK-landscape problem instance with a genetic algorithm. We also found a good basis with a meta genetic algorithm on the problem instance. We changed a basis of the problem instance on the found basis, and could significantly improve the original result of the genetic algorithm.

GECCO '18 Companion, July 15–19, 2018, Kyoto, Japan

ACM ISBN 978-1-4503-5764-7/18/07…$15.00
https://doi.org/10.1145/3205651.3208774

2 A GENETIC ALGORITHM FOR FINDING A GOOD BASIS

The basis can help genetic algorithms create and preserve high-quality schemata. A good basis is such a basis. Changing a basis can be represented by a invertible matrix. We introduce a genetic algorithm to find a good basis with the matrix encoding.

2.1 Elementary Matrices Encoding

It is known that every invertible matrix is represented as a product of elementary matrices [1]. Hence, we represent a solution in $GL_n(\mathbb{Z}_2)$ as a product of elementary matrices. We can consider the representation by a variable-length linear string, of which each element is an elementary matrix [8]. Any recombination for variable-length string can be used.

2.2 Genetic Operator

2.2.1 Crossover Operator. Before recombination, we optimally aligns two strings. The crossover is performed after an optimal alignment with minimal Hamming distance by interleaving "—" anywhere in the strings. The offspring generated by uniform crossover applied to aligned parents after removing "—". The optimal alignment of the two strings is obtained by the Wagner-Fischer algorithm [4].

2.2.2 Selection, Mutation, Replacement. The selection operator applies tournament selection by choosing two parents. We make a good individual to choose four times more often. The mutation operator applies one of the following three operations—insertion, deletion, or replacement—to each string according to a probability. The replacement operator applies the preselection proposed by Cavicchio. This replaces the poor quality solution of the two parent solutions.

2.2.3 Fitness. We use a meta genetic algorithm to find a good basis for target problem. We define the fitness of a basis as follows: the objective of the best solution that is obtained by the target problem with the changed basis.

3 EXPERIMENTS

The experiments deals with the NK-landscape model. The NK-landscape model [2] is a mathematical model described by Kauffman as a tunably rugged fitness landscape. When $K = 0$, the fitness landscape is very smooth. However, the landscape becomes rugged as K approaches N.

3.1 NK-landscape Model

The NK-landscape model consists of a string of length N and fitness is contributed to each character. In addition, the value for each character is set depending on the other K characters. These are known as fitness contributions; fitness values are set depending on one character and other K characters. These fitness contributions are often randomly chosen from a particular probability distribution.

The structure of the experiment is as follows. Initialize fitness contributions using the uniform distribution on interval $[0, 1]$. Calculate the fitness function as the average of the fitness contributed to each character. We find the solution of NK-landscape problem using a genetic algorithm. The algorithm is steady-state genetic algorithm with tournament selection, one-point crossover, preselection replacement. The following shows the experimental results of the NK-landscape model with generations is $6,000$.

Table 1: Results for NK-landscape model with $N = 100$ according to K

K	Best	Average	SD[1]
3	0.71	0.67	0.02
5	0.69	0.66	0.02
10	0.66	0.58	0.04
20	0.61	0.53	0.04
25	0.63	0.54	0.04

[1] Standard deviation

This shows that the higher the K value, the more complicated the problem of the NK-landscape model.

3.2 Finding a Good Basis on NK-landscape Model

We use a meta genetic algorithm to find a good basis described in Section 2 for an NK-landscape problem instance. We changed the basis of the problem instance. Table 2 shows that the found basis improves the original result.

Table 2: Results of our genetic algorithm for finding a good basis in NK-landscape model with $N = 100$ according to K

K	Best	Average	SD
3	0.76	0.75	0.01
5	0.76	0.76	0.02
10	0.74	0.73	0.02
20	0.68	0.68	0.01
25	0.67	0.67	0.01

Figure 1 shows that the optimal solution and the NK-landscape model was obtained through the basis search with $N = 100$ and $K = 10$.

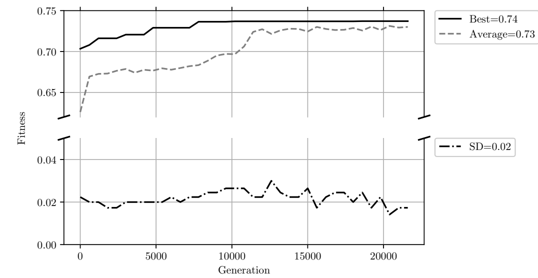

Figure 1: Results of our genetic algorithm for finding a good basis on NK-landscape model with $N = 100$ and $K = 10$

4 CONCLUSIONS

It was not easy to determine a good basis from the NK-landscape model; in addition, we determined how useful it was. Currently, it is costly to find a good basis; However, if there is a good basis, it is sure to obtain a better-quality solution.

Searching for a solution through a genetic algorithm to determine whether the basis is an appropriate basis of the problem space. This is not practical owing to its time cost. However, it may be possible to practically apply a basis to a genetic algorithm if a method of heuristically evaluating the basis is devised. For example, while we can evaluate the basis with epistasis, it is realistically impossible to obtain the value of epistasis between genes. A good basis can therefore be obtained by estimating the basis in the direction of decreasing epistasis as the good one by using the estimate of epistasis through sampling [6, 9].

ACKNOWLEDGMENTS

This research was supported by a grant [KCG-01-2017-05] through the Disaster and Safety Management Institute funded by Korea Coast Guard of Korean government.

REFERENCES

[1] Stephen H. Friedberg, Arnold J. Insel, and Lawrence E. Spence. 2003. *Linear Algebra, 4th Edition.* Pearson.

[2] Stuart A. Kauffman and Edward D. Weinberger. 1989. The NK Model of Rugged Fitness Landscapes and its Application to Maturation of the Immune Response. *Journal of Theoretical Biology* 141, 2 (1989), 211–245.

[3] Yong-Hyuk Kim and Yourim Yoon. 2008. Effect of Changing the Basis in Genetic Algorithms using Binary Encoding. *KSII Transactions on Internet and Information Systems* 2, 4 (2008), 184–193.

[4] Gonzalo Navarro. 2001. A Guided Tour to Approximate String Matching. *Comput. Surveys* 33, 1 (2001), 31–88.

[5] Eduardo Rodriguez-Tello and Jose Torres-Jimenez. 2004. Improving the Performance of a Genetic Algorithm Using a Variable-Reordering Algorithm. In *Proceedings of Genetic and Evolutionary Computation Conference.* Springer, 102–113.

[6] Dong-Il Seo, Yong-Hyuk Kim, and Byung Ro Moon. 2003. New Entropy-Based Measures of Gene Significance and Epistasis. In *Proceedings of the Genetic and Evolutionary Computation Conference.* Springer, 1345–1356.

[7] Danica Wyatt and Hod Lipson. 2003. Finding Building Blocks through Eigenstructure Adaptation. In *Proceedings of the Genetic and Evolutionary Computation Conference.* Springer, 1518–1529.

[8] Yourim Yoon and Yong-Hyuk Kim. 2014. A Mathematical Design of Genetic Operators on $GL_n(\mathbb{Z}_2)$. *Mathematical Problems in Engineering* 2014 (2014). Article ID 540936, 8 pages.

[9] Dong-Pil Yu and Yong-Hyuk Kim. 2017. Relationship between Epistasis on Performance of Genetic Algorithm. In *Convergence Research Letter*, Vol. 3. The Convergent Research Society among Humanities, Sociology, Science, and Technology, 675–680.