Function-discovery-system by the evolutional strategy using the search-accumulation

Mitsutoshi SAITO and Seiichi SERIKAWA
Graduate School of Electric Engineering
Kyushu Institute of Technology
1-1 Sensui-cho, Tobata-ku, Kitakyushu-shi, Fukuoka, 804-8550, JAPAN

Abstract

Recently, the system using bug type of artificial life was proposed for discovering the function, and it was improved. It is one of the extended models of Genetic Algorithm (GA) and Genetic Programming (GP). However, the function is not found sometimes when the observation data is very complicated.

In this study, a new concept is introduced so that the function-search can be applicable to the complicated observation data. The function-search by S-System is executed two or more times as a basic idea. This is called "search-accumulation". To confirm the validity of search-accumulation, equal-loudness-level contours (ISO226) are used as the observation data. Since the contours are very complicated, it is very difficult to express the counters as a function. In reality, the function-search was tried by the conventional S-System, but the function does not agree well with the data. By the use of the strategy of search-accumulation, the function in agreement well with the contours is found successfully. Thus, the validity is conformed.

Keywords: search accumulation, genetic programming

1. Introduction

Recently, a variety of problem-solving systems that use genetic algorithm (GA) and genetic programming (GP) are proposed. The purpose of GA is mainly the optimization of the numerical value, while that of GP is the generation of the symbol. Koza proposed the function-discovery-system using GP [1]. It generates the approximate function that shows the relationship between input data and output data in observation data. In GP, the chromosome which represents a function is expressed by the tree structure. It is optimized by the crossover and mutation. However, there are problems that the schema is destroyed and the length of generated function becomes extremely long. To solve the problem, the function-discovery-system that used the concept of an artificial life was proposed [2]. This is called "S-System", because it is the system that uses the concept of sexual and asexual reproduction. The improved model was also proposed [3]. However, both GP and S-System cannot discover the approximate function, when the observation data is complex.

In this study, a new concept is introduced to generate the approximate function corresponding with the complicated data. A basic idea is the execution of the function-search by S-System two or more times. This strategy is called "search-accumulation" in this study. As the result, the generated function agrees well with the complicated observation data.

2. The model for discovering function

2.1 Summary of S-System

The S-System is one of the evolutional strategies, and the mail flow is the same as that of GA and GP. The chromosome and the fitness are designed according to the purpose of applied field. In general, the fitness is defined as an evaluation-function in which the quality of the chromosome is expressed. The main flow of evolutional strategy is as follows.

(i) A lot of chromosomes are generated at random as the first stage. (ii) The chromosome is selected. The selected probability of the chromosome with high fitness is high than of the chromosome with low fitness. (iii) The crossover and the mutation are practiced between two selected chromosomes. As a result, a new chromosome is born. (iv) The procedures of (ii) and (iii) are repeated until the fitness exceeds a certain threshold value.

As the result, the chromosome changes and the value of fitness increases, according as the generation proceeds.

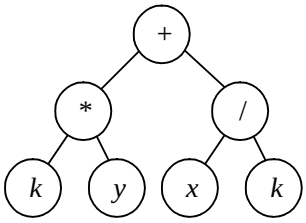

Fig.1 The example of chromosome for GP (tree structure of $k \cdot y + x / k$).

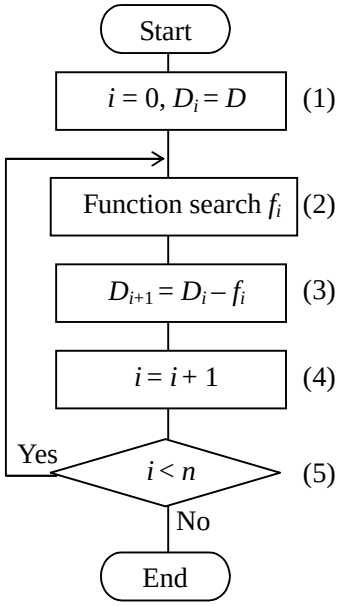

Fig. 2 The flowchart of function-discovery-system using search-accumulation.

The length of chromosome for GA is fixed, but that for GP changes. The purpose of GA is the optimum of the numerical value. On the other hand, that of GP is the generation of the symbols. The tree structure is regarded as the chromosome for GP as shown in Fig. 1. The S-System also represents the tree structure as the chromosome. The crossover of GP is different from that of S-System. For GP, it is practiced between the chromosomes with different tree structure. Therefore, the form of tree structure dramatically changes. For example, when the chromosome shows a function as shown in Fig.1, the form of function greatly changes by the crossover. Thus, the solution is changeable. On the other hand, the function form doesn't change by the crossover in S-System. The concepts of homogeneity and heterogeneity have been introduced into S-System. If two individuals have the same tree structure, they are called homogeneity. The individual with a different tree structure is called heterogeneity. The crossover is practiced between homogeneities. It is called “sexual reproduction”. The tree structure is not broken by the sexual reproduction, but only the constant value included in the function changes. This is because the tree structure of homogeneity is equal. Therefore, the constant value is optimized as the generation proceeds. On the other hand, the individual without homogeneity generates its two copies. This is called “asexual reproduction”. By the strategy, the sexual reproduction can be practiced to the next generation, because two homogeneities are born by the asexual reproduction. In S-System, the change of tree structure occurs at the mutation. As the result, the optimization of the numerical value is practiced, the length of function doesn't become long compared with GP, and the solution is stabilized. For more details, refer to Ref [2].

2.2Function-discovery by search-accumulation

In this study, a new concept is introduced so that the function-search can be applicable to the complicated observation data. The function-search by S-System is executed two or more times. This is called “search-accumulation”. Here, the observation data is assumed to be D, and its approximate function is assumed to be f. Figure 2 shows the flowchart of the function-discovery-system using search-accumulation. The details are described as follows.

(1) The counter i is set to be 0, and observation data D is regarded as the initial data D_0.
(2) An approximate function that agrees with the observation data D_i is searched by S-System. The obtained function is assumed to be f_i.
(3) The difference (i.e. D_i - f_i) between D_i and f_i is regarded as the new observation data D_{i+1}.
(4) One is added to the value of the counter i.
(5) The search finishes when the value of counter i reaches a threshold. Otherwise, jump to (2).

After the search, the approximate function f is obtained by the following equation.

$$f = \sum_{i=0}^{n} f_i \qquad (1)$$

When n = 0 in Eq. (1), the equation corresponds to the function discovered by the conventional S-System. Thus, the proposed search-accumulation is regarded as the extended model of conventional S-System.

3. Experimental Result and investigation

To confirm the validity of the proposal method, equal-loudness-level contours (ISO226) [4] are used as an observation data. From the data, good approximate

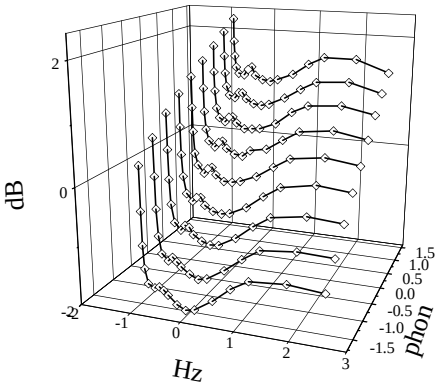

Fig.3 Equal-loudness-level contours about sound (ISO226) [4].

Table 1. The value of parameters for S-System.

Number of chromosomes	900
Maximum generation	4000
Generation gap	0.5
Number of homogeneities at next generation	20
Number of gates at next generation	200
Tournament size	2
Crossover rate	0.2
Limitation of the number of homogeneities	30
Maximum number of gates	300
Threshold for going back	0.01
Mutation rate of virus	0.3
Mutation rate of bugs	0.2
Mutation rate of constant	0.2
Maximum depth of tree	8
Threshold of fitness for ending program	1

function was not obtained by the conventional S-System.

Figure 3 shows the contours. The loudness level [dB] that man psychologically perceives changes by frequency [Hz] and the sound pressure level [phone]. In another word, the loudness [dB] is represented as a function *f* of the frequency [Hz] and the sound pressure level [phone]. In this study, the frequency [Hz] and the sound pressure level [phone] are regarded as the input data, and the loudness level [dB] is regarded as the output data. The approximate function *f*(Hz, phone) is obtained from those data. The search-parameter of S-System is shown in Table 1. The search is tried under

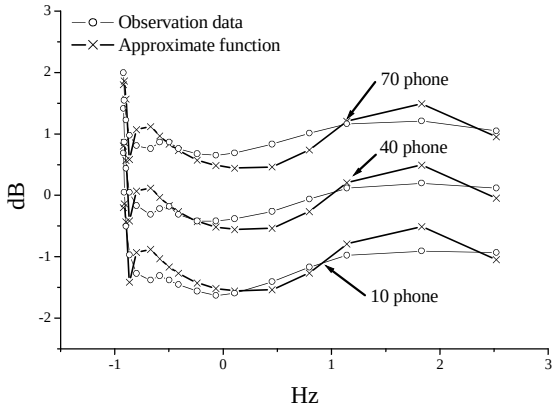

Fig.4 Example of approximate function by Condition I (fitness=0.781).

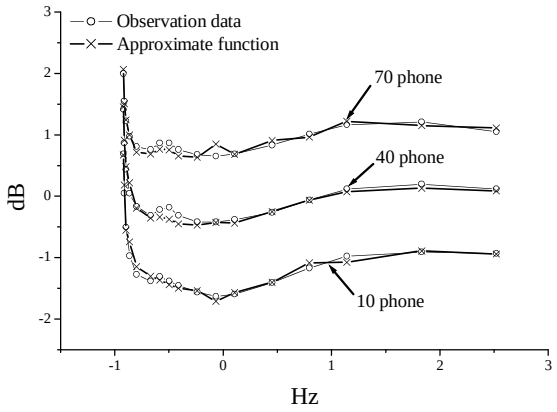

Fig.5 Example of approximate function by Condition II (fitness=0.918).

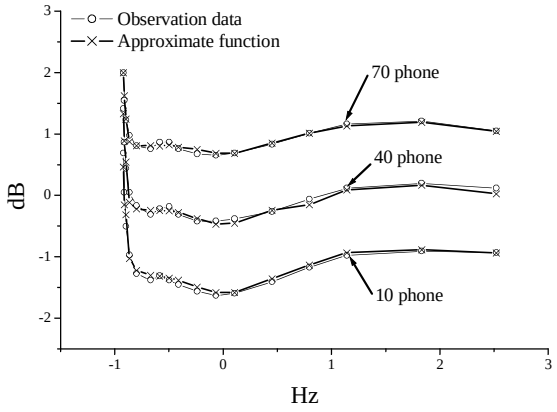

Fig.6 Example of approximate function by Condition III (fitness=0.927).

the following three conditions.

Condition I: Conventional method.

Condition II: Search-accumulation I.

Condition III: Search-accumulation II.

As for Condition II and Condition III, the method of the accumulation is different. In Condition II, after the

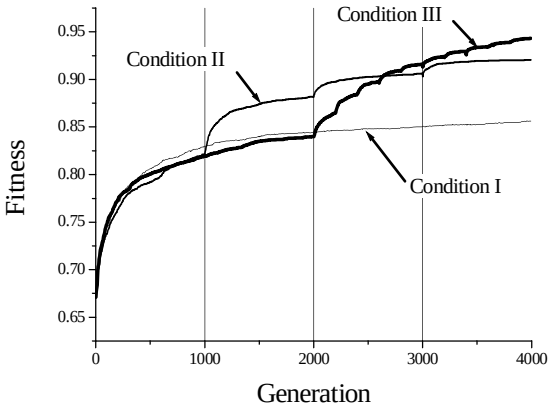

Fig. 7 Relationship between the generation and the fitness in Condition A, B, and C (average of 30 trials).

search of 1000th generation, the accumulation is practiced. It is repeated every 1000 generation. Total generation is 4000 ($=1000\times4$). In Condition III, the first search is practiced until the 2000th generation. After that, the accumulation is repeated every 200 generation. Total generation is 4000 ($=2000+200\times10$). The examples of the experimental result are shown in Fig. 4, 5, and 6. To avoid the confusion of drawing curves, only the results of 10, 40, and 70 phones are shown. As understood from these Figs, the waveform by Condition II and Condition III are corresponding with the observation data well compared with that of Condition I. Figure 7 shows the relationship between the generation and the fitness in Condition I, II, and III. There is little increase in fitness, when the generation exceeds 2000 for Condition I. As for Condition II and III, it is understood that the fitness has improved whenever the search-accumulation is repeated. Therefore, the validity of the proposal method is mentioned.

The fitness rapidly increases in Condition II after the 1000th generation. This is corresponding to the generation when the first search-accumulation started. On Condition III, the fitness rapidly increases after the 2000th generation is over. This is also corresponding to the generation when the first search-accumulation started. Thus, it is understood that fitness increases by the search-accumulation. The final fitness of Condition III is higher than that of condition II. The frequency of the search-accumulation of condition III is more than that of condition II. From the above-mentioned, even if the number of total generations is the same, it is understood that the search-ability is different if the accumulation-frequency is different.

4. Conclusions

The search-accumulation of the function-discovery by S-System has been proposed in this study. When this method is used, the function corresponding well with the complicated observation data is generated. To confirm the validity of the proposed method, equal-loudness-level contours (ISO226) are used as an observation data. In conventional method, the function in agreement with this data is not obtained. On the other hand, the function corresponding well with the observation data has been discovered by proposed method.

As the problem in the future, there is the reduction of the length of discovered function. In addition, the effectiveness should be verified by applying to more practical problems.

References

[1] J. Koza, *Genetic Programming II*, *Auto Discovery of Reusable Subprograms*, *MIT Press*, p.109, (1994).

[2] S. Serikawa and T. Shimomura, “Proposal of a System of Function-Discovery Using a Bug Type of Artificial Life”, *Trans. IEE*, Vol.118-C, 2, p.170, (1998).

[3] S. Serikawa and T. Shimomura,“Improvement of the Search Ability of S-System (A Function-Discovery System)”, *Trans. IEE*, Vol.120-C, 8, p.1281, (2000).

[4] National Astronomical Observatory, *Chronological Scientific Tables 2006, Maruzen Co., Ltd*, p.424, (2005).