# RESEARCH ON AN IMPROVED GENETIC ALGORITHM BASED KNOWLEDGE ACQUISITION

**LI-MIN SU, HONG ZHANG, CHAO-ZHEN HOU, XIU-QIN PAN**

Dept. of Automatic Control, Beijing Institute of Technology, Beijing, China.
E-MAIL: sulimin2000@163.com

**Abstract:**

**Knowledge acquisition is the "bottleneck" of the building of an expert system. In this paper, based on the optimization model, an improved genetic algorithm applied to knowledge acquisition of a network fault diagnostic expert system is described. The algorithm applies operators such as selection, crossover and mutation to evolve an initial population of diagnostic rules. Especially, a self-adaptive method is put forward to regulate the crossover rate and mutation rate. In the end, a knowledge acquisition problem of a simple network fault diagnostic expert system is simulated, the result of simulation shows that the improved approach can solve the problem of convergence better.**

**Keywords:**

**Expert system; Knowledge acquisition; Fault diagnosis; Genetic algorithm**

## 1 Introduction

Expert system technology has been widely used to solve the network fault diagnostic problem. Nonetheless there is a specialised problem that the network fault diagnostic expert system came across-- knowledge acquisition. Expert knowledge is good or bad effect the performance of the whole system directly [1]. Now there are two methods to solve the knowledge acquisition, machine learning automatically and engineer acquiring from domain experts. In most condition, it is difficult to the domain experts to express their experience clearly, so the efficiency of the later method is lower. For this reason, an improved genetic algorithm is introduced from the point of view of optimization in this paper; it can solve the problem of knowledge acquisition in a certain degree.

The remainder of this paper is composed of four sections. In section two, we give a brief description of traditional genetic algorithm that is needed for an understanding of the paper. Section three describes a machine learning method based on the improved genetic algorithm. In section four, the knowledge acquisition problem of a simple network fault diagnostic expert system is simulated and the simulation results are presented. The last section concludes with a summary of the paper's key messages.

## 2 Genetic Algorithm

The genetic algorithm(GA) is a machine learning technique. It was invented based on the principles of genetic variation and natural selection [2]. In recent years, GA has become increasingly popular as a method for solving complex search problems in a large number of different disciplines. The appeal of GA comes from its simplicity and elegance as algorithms as well as from their power to discover good solutions rapidly for difficult high-dimensional problems.

In their original fashion, the genetic algorithms operate on population of so-called chromosomes. These are binary strings that represent possible solutions in a certain way [3]. During the evolution process, chromosomes are evaluated and the fitness values associated with chromosomes are computed. The chromosomes are selected from the population according to the fitness distribution. Moreover, Genetic operators such as crossover (exchanging substrings of two parents to obtain two descendant) and mutation (flipping individual bits) are then applied probabilistically to the population to produce a new generation of individuals. The GA is considered to be successful if a population of highly fit individuals evolves as a result of iterating this procedure.

## 3 Machine learning based on improved genetic algorithm

### 3.1 Definition of indicator function

In a fault diagnostic expert system, the knowledge can be expressed in the form of rules. For example:

$$R\text{: If } r_1 \& r_2 \& \cdots \& r_m \text{ Then} \tag{1}$$

In formula (1), $r_i (i=1,2, \cdots,m)$stands for the precondition of the rule corresponding to the symptom of the system failure, and $C$ stands for the conclusion of the rule corresponding to the type of the failure. The purpose of the algorithm is to find an optimal rule in the rule space, and make it in great similitude of its proper example.

The precondition can be expressed as follows:

$$R=(r_1, r_2, \cdots,r_m) \tag{2}$$

In the genetic algorithm, chromosome is expressed with binary strings, number 1 means that the symptom is appearance, and the meaning of number 0 is in opposition.

So the value of $r_i(i=1,2,\cdots,m)$has only two choices, number 1 or 0. Now we let $PE$ the set of $R$'s proper example and $CE$ the set of its counterexample.

During the process of machine learning, a performance function to evaluate the set of rules is needed, and it directs the algorithm to perform in the anticipant direction. Suppose $PE$ is a $l$-dimensional vector($PE_i(i=1,2,\cdots,l)$), $CE$ is a $s$-dimensional vector($CE_i$ $(i=1,2,\cdots,s)$ ), then they can be expressed as follows:

$$PE_i = (r_{i1}, r_{i2}, \ldots, r_{im}) \quad i = 1,2,\ldots,l \tag{3}$$

$$CE_i = (\overline{r_{i1}}, \overline{r_{i2}}, \ldots, \overline{r_{im}}) \quad i = 1,2,\ldots,s \tag{4}$$

A random rule $R'$ in rule space can be expressed as formula (5).

$$R': \text{ If } a_1 \& a_2 \& \ldots \& a_m \text{ Then } C' \tag{5}$$

The similar degree between $R'$ and its proper example or counterexample can be expressed with a cosine value between two corresponding vectors. The formulae were showed as follows [4]:

$$r(R', PE_i) = \frac{\sum_{j=1}^{m} a_j r_{ij}}{\sqrt{\sum_{j=1}^{m} a_j^2}\sqrt{\sum_{j=1}^{m} r_{ij}^2}} \quad i = 1,2,\ldots,l \tag{6}$$

$$r(R', CE_i) = \frac{\sum_{j=1}^{m} a_j \overline{r_{ij}}}{\sqrt{\sum_{j=1}^{m} a_j^2}\sqrt{\sum_{j=1}^{m} (\overline{r_{ij}})^2}} \quad i = 1,2,\ldots,s \tag{7}$$

$r(R',PE_i)$ stands for the similar degree between rule $R'$ and its proper example, $r(R',CE_i)$ stands for the similar degree between rule $R'$ and its counterexample, $a_i(i=1,2,\ldots,m)$ is the precondition of a random rule in the rule space, the meaning of $r_{ij}$ and $\overline{r_{ij}}$ are expressed in formulae (3) and (4).

Then the problem of machine learning can be converted into an optimal problem, i.e. finding a rule like formula (5) in the rule space and making the value of $r(R',PE_i)$ and $r(R',CE_i)$ reach maximum and minimum respectively. Aiming at the optimum target, using the min-max rule, an optimum indicator function $f(x)$ is presented. It can be expressed as follow:

$$\max f(x) = \frac{1 + \min\{r(R', PE_i), i = 1,2,\ldots,l\}}{1 + \max\{r(R', CE_i), i = 1,2,\ldots,s\}} \tag{8}$$

### 3.2 Genetic operator

Genetic operators are used to alter the composition of chromosomes. The fundamental genetic operators, selection, crossover, mutation and modification are used to create children (or individuals in the next generation) that differ from their parents (or individuals in the previous generation).

There are many methods to select two parents from the old population, and different genetic algorithm methods can be obtained by using different selection methods. In this paper, the Roulette wheel selection method [5] is employed. This method is essentially a weighted random selection approach. Suppose the sum of the fitness values of all chromosomes within the population is $F$, the probability that a chromosome with fitness value $f_i$ is selected is $p$, then

$$p(f_i) = f_i / F = f_i / \sum_{i=1}^{m} f_i \tag{9}$$

During the process of evolution, crossover and mutation are the most important operators. So the selection of crossover rate $p_c$ and mutation rate $p_m$ is also important, it affects the astringency of approach directly. In traditional GA, the values of $p_m$ and $p_c$ are constant. Then if the value of $p_c$ is bigger, the approach will be unstable, otherwise $p_c$ is smaller, the mature convergence will not be achieved. A certain mutation rate $p_m$ also can avoid the approach to be get in part optimal. So in this paper we adopt a self-adaptive regulation method, the values of $p_c$ and $p_m$ can be regulated correspondingly with the variation of fitness values. The definition formulae of $p_c$ and $p_m$ are expressed as follows:

$$p_c = \begin{cases} k_c - (k_c - k_{c0}) \dfrac{(f_{\max} - f_c)}{(f_{\max} - f)} & f_c \geq f \\ k_c & f_c < f \end{cases} \tag{10}$$

$$p_m = \begin{cases} k_m - (k_m - k_{m0}) \dfrac{(f_{\max} - f_m)}{(f_{\max} - f)} & f_m \geq f \\ k_m & f_m < f \end{cases} \tag{11}$$

In the formulae above, $k_c$=0.95, $k_{c0}$=0.6, $k_m$=0.15, $k_{m0}$=0.001, $f_c$ is the bigger fitness value between two individuals to be corssed, $f_m$ is the fitness value of the individual to be mutated, $f_{\max}$ is the biggest fitness value in population, and $f$ is the average fitness value.

Using the self-adaptive regulation method can improve the crossover and mutation rates of excellent individuals accordingly. After crossing and mutating, the descendants not only inherit their parents' better characteristics, but also add the new characteristics through mutation.

For the complicated fault diagnosis problem, multiple solutions may exist. In order to find all reasonable solutions, modification operator is adopted. At the first generation, we pick a copy for each of the best solutions from the population and store them in a specially designed array. In each of the follow-up generations, we check if the best solutions stored in the current population are better than the solutions in the array. If it is true, we use the best solutions in the current generation to replace the record of array. Otherwise we do not change them. This can ensure the continuity of population, and avoid the elimination of good solution.

## 4 Simulation experiment

### 4.1 Acquisition of diagnostic examples

In order to test the validity of the improved genetic algorithms, we took three simple network failures, protocol error, network management station(NMS) down and router broken as example. Through calling MIB variables from interface group, typical diagnostic example can be acquired. Table 1 shows part of the failure examples; table 2 shows the corresponding relationship between symbols and symptoms.

Table 1. Fault example

| Fault Number | Fault Type | Fault Example |
|---|---|---|
| 1 | Protocol Error | $(s_1, s_3, s_5)(s_1, s_2, s_7)$ $(s_1, s_3, s_8)(s_1, s_6, s_9)$... |
| 2 | NMS Down | $(s_3, s_4, s_7)(s_3, s_5, s_7, s_{10})(s_3, s_6, s_7)$... |
| 3 | Router Broken | $(s_3, s_5, s_6)(s_3, s_8, )$ $(s_3, s_6, s_9)(s_3, s_4, s_8)$... |

Table 2. symptom of the fault

| Symbol | Symptom |
|---|---|
| $s_1$ | Protocol type of interface is unmatched |
| $s_2$ | Managing state of interface is closed |
| $s_3$ | Current state of interface is closed |
| $s_4$ | The number of wrong packet is increased |
| $s_5$ | Average flux of input is increased |
| $s_6$ | Average Flux of output is increased |
| $s_7$ | DCE down |
| $s_8$ | The number of resetting is increased |
| $s_9$ | Wrong CRC checkout |
| $s_{10}$ | DTE down |

### 4.2 Approach realization

In the following simulation, the settings of the GA parameters are $k_c$=0.95, $k_{c0}$=0.6, $k_m$=0.15, $k_{m0}$=0.001, the population size=10, the maximum permitted iteration step=150, and the stop criterion is that the maximum permitted iteration steps has been reached. Figure 1summarizes the basic steps of the improved algorithm.

### 4.3 Simulation results

Through the iterative operation above, three chromosomes with highest fitness values are obtained. Decode the learning results, then three diagnostic rules can be expressed as follows:

(1) IF ($p_1$=1) THEN Protocol Error
(2) IF ($p_3$=1)&($p_7$=1) THEN NMS Down
(3) IF ($p_2$=0)&($p_3$=1) THEN Router Broken

```
begin
code the fault example with binary strings;
initialize the population;
calculates each individual's fitness;
//use formulae (6), (7)and (8)
do
{
selected operation;        //use formula (9)
self-adaptive crossover;   // use formula (10)
self-adaptive mutation;    // use formula (11)
calculates the fitness values;
replace the bad result with the best one;
evaluate the children's fitness;
}
while the stopping criterion is achieved.
```

Fig.1. basic steps of the improved algorithm

In order to validate the performance of improved genetic algorithm, traditional genetic algorithm is also performed in the same condition. Table 3 compares the learning results of two approaches.

Table 3. Comparision of results

| Algorithm | Parameter setting | $N_{gen}$ (time) | $T_{con}$ (s) | $f_{stb}$ |
|---|---|---|---|---|
| General GA | $p_c$=0.9 $p_m$=0.1 | 125 | 14.72 | 1.2 |
| Improved GA | $k_c$=0.95, $k_{c0}$=0.6 *km=0.15* *km0=0.001* | 22 | 2.21 | 1.64 |

In table 3, $N_{gen}$ is the iteration step number, $T_{con}$ is the convergence time, $f_{stb}$ is the final stable fitness value. From Table 3, we can see that the improved GA can achieve stabilization rapidly. The convergent speed of improved GA is faster than that of traditional GA, and the stable fitness value of improved GA is bigger than that of traditional GA.

## 5 Summary

Based on the improved GA, an intelligent knowledge acquisition method of fault diagnostic expert system is introduced in this paper. It can solve the problem of knowledge learning in the fault diagnostic expert system effectively, and exert its advantage in the complicated diagnostic system with great information. The approach can improve the GA's convergence capability greatly, and the rules acquired from it can reflect the real failure correctly.

Nonetheless the final fitness value of our approach is still lower, maybe it would be better through regulating the other parameters of GA or improving the formulae of genetic operation, the line of research for future work is that how to enhance the fitness value greatly.

## References


[1] Todd, E. Marques. A Symptom-Driven Expert System for Isolating and Correcting Network Faults. IEEE Communications Magazine, March 1988, Vol. 26, No. 3

[2] M. Mitchell. An Introduction to Genetic Algorithm. MIT Press, Cambridge, MA, 1996

[3] Mark Jelasity, Jozsef Dombi. GAS, a concept on modeling species in genetic algorithms. Artificial Intelligence, 1998(99), pp.1-19

[4] Zang Liang, Xia Delin, Yan Puliu. Knowledge Acquisition and Its Application in Intelligent Fault Diagnosis of Network. Computer Engineering, 2001(10): 53-54 (in Chinese)

[5] Benjamin Han, Shie-Jue Lee. A genetic algorithm approach to measurement prescription in fault diagnosis. Iformation Sciences, 1999(120), pp.223-237