RESEARCH ON THE APPLICATION OF IMPROVED SHUFFLED FROG LEAPING ALGORITHM IN MECHANICAL FAULT DIAGNOSIS

Qingjun Guo

1 Chongqing Jianzhu College, Chongqing 400072, China
E-mail: 672139860@qq.com

ABSTRACT: For some difficult problems in the process of mechanical equipment fault diagnosis, this study applies some key theories on the widely applied swarm intelligence algorithm- Shuffled Frog Leaping Algorithm and improved genetic algorithm in the diagnosis of mechanical failure, starting from "intelligent optimization", the study realizes intelligent solution to cost function and the number of clusters in machine learning of optimal sensor placement and parameter optimization of neural network model in diagnostic system. The research results confirm the local convergence and global convergence of the frog individual updating model. Combining the Markov model and the theories related to expected convergence time of the algorithm, the study carries out the theoretical analysis of the convergence speed of SFLA and the analysis of the complexity of the algorithm itself, making the theory of the SFLA algorithm perfect. The study summarizes the law of the influence of parameter setting on the performance of the algorithm, and realizes the accurate classification and pattern recognition of unsupervised data. The research results are helpful to the accuracy of mechanical equipment fault diagnosis.

KEYWORDS: fault diagnosis, shuffled frog leaping algorithm, optimal sensor placement, SFLA

1 INTRODUCTION

Mechanical fault diagnosis is very important science to identify the running state of machine or units. Its core work is to facilitate the acquisition transmission, processing, regeneration and utilization of diagnostic information, and thus can accurately identify and handle mechanical problems, guarantee normal operation of equipment and ensure safe production (Li et al., 2016; Zhao et al, 2016). In recent years, the mechanical equipment is becoming larger and larger, and the working situation is becoming more and more complicated. How to easily and quickly extract useful function information from the equipment, judge the running state of the equipment and make accurate diagnosis of the fault will be new challenges to the mechanical equipment fault diagnosis at present.

Mechanical fault diagnosis technology is science and technology to monitor and diagnose the running state of mechanical equipment and guarantee the safe operation of mechanical equipment (Wang et al, 2015). With the continuous growth of economy and the constant progress of science and technology, the production equipment used in all walks of life related to national economy and people's livelihood has been developing in the direction of large-scale, high-speed, integration and automation (Sun and Zhao, 2011; Jiang, 2016). Intelligent production of large-scale shield tunneling equipment used in large-scale tunnel construction projects, automation equipment of production lines, large-scale integrated circuit production and manufacturing equipment and aerospace vehicles can greatly save manpower and material resources, and can also improve production efficiency. The industrial revolution has brought all kinds of benefits to people's production and life, and makes it convenient and convenient in all kinds of work (Yang et al, 2013; An, 2017). However, the corresponding problems also appear. As the mechanization degree and composition of large-scale equipment become more and more complicated, and the inter-relation between equipment becomes closer and closer, the requirements for the function of key parts of equipment is very high. Once the key parts of mechanical equipment fail, there will often be chain reaction such as the paralysis of the whole system, which brings irreparable losses to production. Large-scale accidents caused by failure of production equipment at home and abroad often occur, which brings great losses to production and life safety (Snyman, 2000; Chen and Guo, 2015; Li et al., 2016).
The current mechanical fault diagnosis technology has some serious problems such as unclear recognition of fault mechanism, unreasonable fault diagnosis method and relatively weak intelligent diagnosis technology. For the above problems, the study deeply analyzes the theoretical composition and calculation method of the new intelligent algorithm—Shuffled Frog Leaping Algorithm (SFLA), and proposes the mechanical fault diagnosis and analysis technology based on WSFLA, which is applied in practical engineering with reasonable and scientific test method.

2 THEORETICAL MODEL OF OPTIMAL SENSOR PLACEMENT FOR MECHANICAL EQUIPMENT

Assuming that the operating system has n states or fault pattern (fault sources), the set is $FM = \{f_{m_1}, f_{m_2}, \ldots, f_{m_n}\}$, and the initial position of the sensors arranged in the mechanical equipment should be convenient to obtain the relevant fault information of the fault concentration in the system as much as possible. There are m sensors in the system as alternatives, and the sensor at the measuring point i of the initial sensor group is $s_i$, then the set of sensors is recorded as $S = \{s_1, s_2, \ldots, s_m\}$, and the fault state of the system-sensor correlation matrix is marked as a 0-1 type matrix, $\Phi = [\phi_{ij}]$, as shown in Equation (1) (Liang et al., 2017):

$$
\begin{bmatrix}
1 & 1 & 1 & \phi_{1j} \\
1 & 1 & 1 & \phi_{2j} \\
0 & 0 & 0 & \phi_{3j} \\
M & M & M & M \\
\phi_{1i} & \phi_{2i} & \phi_{3i} & \phi_{nj}
\end{bmatrix}
$$

When $i=1, 2, 3 \ldots m$, and $j=1, 2, 3 \ldots n$, the rows in the above matrix represent fault source patterns that each sensor can detect, and the columns represent situations where each fault can be detected by all sensors (Jia et al., 2016). The correlation matrix $\Phi$ describes the relationship between the set of fault mode (FM) and the set of sensors (S), which is the basis of optimizing the sensor configuration for the complex system. $\Phi_{ij}$ is a relation factor with Boolean taken. When $\Phi_{ij}=1$, it indicates that the fault mode $f_{m_j}$ can be detected by the sensor at the measuring point $i$; when $\Phi_{ij}=0$, it indicates that no matter how many sensors are added at the measuring point $i$, no fault mode can be detected (Nikut et al., 2012; Li et al., 2015; Huang et al., 2017).

For the matrix $\Phi$, when a row is all 0, it means that the sensors cannot detect the fault problem at this position no matter how many sensors are added here, and the detection position needs to be adjusted: when rows are all 1, it indicates that the fault sources of the faults in these positions can all be detected, and there are redundant sensors. Thus, the sensors should be reduced (Banerjee et al., 2017; Wang et al., 2015).

Define a 0-1 discrete column vector: $X = [x_1, x_2, \ldots, x_n]$ as candidate solution, $x_i = 1$ indicates that the $i^{th}$ sensor position is selected, and if this position does not require a sensor, $x_i = 0$; Define the cost of installing each sensor as $C_i$ (Lee et al., 2013), the expression of the least cost is:

$$
C = \min \left( \sum_{i=1}^{n} x_i \cdot c_i \right)
$$

In the actual layout process of the project, it is also necessary to take into account various constraints, usually with the number of sensors as constraints; in order to ensure that all fault problems can be detected, it is necessary to set the constraint condition to be that all faults can be detected (Beskos et al., 2013); it is also necessary to consider the effects of a possible sensor failure, when the failure detection rate, isolation rate, and system false alarm rate are often used as additional constraints (Schuck et al., 1998).

In order to guarantee the observable rate of the faults, it must be ensured that the dot product of the faults, any column of the fault-sensor correlation matrix $f_{m_j}$ with the set of candidate sensors $X$ is equal to 1 at least, when the dot product is equal to 0, it indicates that it is difficult to observe this failure mode with $X$ (Kennedy & Pendleton, 2000), that’s, the following formula is satisfied:

$$
FO(k) = F_k \cdot X = \sum_{i=1}^{n} x_i \cdot \phi_{ik} \geq 1, k \in FM
$$

3 IMPROVEMENT OF 0-1 DISCRETE SFLA BASED ON GENETIC ALGORITHM

3.1 Analysis of improved algorithm based on crossover and mutation operator

In the research method of SFLA, only the local optimum frogs are used to update the worst frogs, and the information of other frogs is not reasonably used in the population, so that the overall difference of the population is easily reduced, the convergence speed of the calculation is reduced, and the algorithm may fall into local optimality, which leads to precocity (Qiangel & Habib, 2000). According to the characteristics of 0-1 problem, the
study introduces the crossover and mutation operators of genetic algorithm in the local depth search, and the partial modification is carried out according to updated strategy, as shown in Step 1~Step 4.

Step 1: In order to improve the reliable search progress of the algorithm, the global optimal frog $U^*_{gb}$ and the local optimal frog $U^*_b$ are used as two objects of the crossover operation. During the crossover operation, one point or more points are selected randomly, and then their fitness values are calculated respectively, and the frog with the higher fitness value is selected as $U^{*}_{Bw}$ for the updating $U^*_{Bw}$:

$$U^*_{w} = \begin{cases} b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & b_7 & b_8 \\ U^*_{1B} & 1 & 1 & 1 & 1 & 0 & 1 & 0 \\ U^*_{2B} & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ U^*_{3B} & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ U^*_{4B} & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{cases}$$

Figure 1. Solving results of local average frog at q=4

Step 2: If the fitness value of $U^*_{Bw}$ in the calculation result is better than $U^*_b$ and then replaces the latter, the local search process of the present subgroup can be ended; otherwise, the local average value of the q binary frogs of the present subgroup is calculated, as shown in Figure 1. The probability of occurrence of each binary frog for 0 and 1 is calculated, if the number of occurrence of 0 is more (for example, b3 bits), the corresponding bit of $U^*_b$ is 0, and otherwise (for example, b1 bit), it is 1; when the probabilities of 0 and 1 are equal (such as b4 bit), 0 or 1 may be randomly generated, and Figure 1 depicts a calculation result of a 8-bit binary local average frog when q = 4;

Step 3: According to the distance between $U^*_b$ and $U^*_{Bw}$, the site operation of $U^*_b$ is carried out. The greater the distance between the two is, the greater the probability of mutation is, the new binary frog $U^*_b$ after mutation is prepared for updating;

Step 4: Comparing the fitness value of the new frog, if there is an improvement, the new frog replaces $U^*_b$, and if there is no improvement, a binary frog can be randomly generated to replace $U^*_b$.

3.2 Simulation verification

When the crossover and mutation of GA algorithm are introduced into SLFA, the property of global convergence of original sequence does not change, but its convergence speed and accuracy can be improved obviously (Riddle & Baker, 2010). In order to verify the convergence performance and solution accuracy of the above algorithm, the following test function is used for the test.

$$\min F(x) = -9x_1 + 3x_3 + 3x_5 + 12x_1x_4 + 60x_4x_6$$

(4)

The calculation results are shown in Table 1, where $x^*$ represents the optimal solution, and the average convergence performance of the function values corresponding to the optimal solution is shown in Figure 2.

Table 1. Optimization results comparison

<table>
<thead>
<tr>
<th>Intelligent algorithm</th>
<th>Optimum solution ($x^*$)</th>
<th>$F(x^*)$</th>
<th>Iterative steps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>The fastest</td>
</tr>
<tr>
<td>D-SFLA</td>
<td>100110,101001,011010</td>
<td>6</td>
<td>--</td>
</tr>
<tr>
<td>GA</td>
<td>100110,101001,011010</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>ID-SFLA</td>
<td>100110,101001,011010</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>
In this example, the fitness functions of the four intelligent algorithms are directly converted by the objective function. However, GA, D-SFLA and ID-SFLA can accurately find out three exact solutions of this function. D-SFLA and ID-SFLA have the same convergence speed and all exceed GA, but their convergence accuracy is the same. A particular practical problem can have several local extreme points in its definition space. An example shows that a 16-dimensional knapsack problem can have 50-200 extreme points.

4 OPTIMAL PLACEMENT OF SENSORS IN CONDITION MONITORING OF GEARBOX

Gearbox is an essential part of variable speed transmission which is used in all kinds of machinery. Its normal operation can involve with the working state of the whole machine or the whole mechanical system. Its serious problem may bring great catastrophic accidents. Therefore, it’s practically significant to diagnose the fault (Liu et al, 2015). In this study, GA, D-SFLA and ID-SFLA are used to properly optimize the type, quantity and installation position of sensors in gearbox condition monitoring, so that the optimized sensor placement can provide the most valuable information at the lowest cost and improve the performance of condition monitoring and fault diagnosis system (Chen et al, 2014).

4.1 Test scheme

Based on the theory of SFLA, an experimental study on the optimal placement of detection points and sensor types is carried out in the sensor monitoring network system with a certain type of two-stage gearbox as research object.

The set of fault states of the gearbox is expressed as \( FM = \{ fm_1, fm_2, fm_3, fm_4, fm_5, fm_6, fm_7, fm_8, fm_9, fm_{10} \} \), where \( fm_1 \) represents the fracture of gear tooth, \( fm_2 \) is wear of gear tooth, \( fm_3 \) is plastic deformation of gear tooth, \( fm_4 \) is fatigue of gear tooth surface, \( fm_5 \) is gear burn, \( fm_6 \) is gear burn, \( fm_7 \) is the gear shaft, \( fm_8 \) is bearing failure, \( fm_9 \) is serious oil leaking of the lubrication system, and \( fm_{10} \) is oil deterioration of the lubricating system. The occurrence probability of these 10 fault modes is 0.11, 0.182, 0.124, 0.121, 0.081, 0.054, 0.052, 0.21, 0.052 and 0.051 respectively.

4.2 Analysis on application of SFLA and results

Based on actual faults and field experience, combined with sensor types and different installation locations, a fault-sensor \( \Phi \) correlation matrix of the gearbox is established as follows:

\[
\Phi = \begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 0 & 1 & 1 & 1 & 1 & 0
\end{bmatrix}
\]

The improved SFLA is applied to the sensor configuration optimization model, which solves the following problems: the meaning of frog individual, the determination of frog population size and the determination of fitness function.

![Figure 3. Average convergent curves of three algorithms](image)

<table>
<thead>
<tr>
<th>Intelligent algorithm</th>
<th>Iterative steps</th>
<th>Iterative steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-SFLA</td>
<td>1.57</td>
<td>2</td>
</tr>
<tr>
<td>GA</td>
<td>1.89</td>
<td>2.88</td>
</tr>
<tr>
<td>ID-SFLA</td>
<td>1.61</td>
<td>1.58</td>
</tr>
</tbody>
</table>

Table 2. Comparison of optimization results for fitness values
The calculation results of the three algorithms are shown in Table 2. By analyzing the data in Table 2 and Figure 3, it can be seen that the accuracy of GA for solving the problem is low, the error is larger, the solution result is abandoned. The accuracy of the solution obtained by D-SFLA is improved, but the success rate is low, and the stability is poor. According to statistics, only 10 of every 20 independent calculations found a probability of a minimum of about 50%. However, the improved ID-SFLA is better than the other two methods in solution speed and accuracy, and the improved algorithm has better stability. With a further analysis of the curve of Figure 3, it can be found that the fastest solution of ID-SFLA can be completed with $N_p=2$ and the slowest solution can be completed with $N_p=107$. This fully demonstrates the efficiency of the algorithm.

According to the solution result of ID-SFLA, the optimal solution $x^*=\{000101000100100\}^T$, that’s, the set of selected sensors is $\{s_5, s_7, s_9\}$, FDR and FIR of the test system meet the accuracy requirements, as shown in the data from Table 3. The correct distinguishing performance of observability and special fault type can also meet the requirements. The number of selected sensors is obviously less than that of alternative sensors, which reduces the cost of testing and reduces the complexity of system testing. It shows that the improved SFLA, as a new intelligent algorithm, has great effectiveness and superiority in sensor configuration model optimization.

Table 3. Comparison of optimization results

<table>
<thead>
<tr>
<th>System requirements</th>
<th>Sensor set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{reg}$</td>
<td>$I_{reg}$</td>
</tr>
<tr>
<td>≥98%</td>
<td>≥95%</td>
</tr>
</tbody>
</table>

According to the calculation results in this study, the solution of the optimal sensor network placement is $F_{min}=1.6$, which is considered to be the global optimal solution. To overcome the random error, totally 20 times of independent calculation is carried out for each parameter combination and its success rate is the percentage of the number of times that the global optimal solution is successfully found in these 20 times of independent calculation. The average value the fastest value $W$ and the maximum value of global iterations are the statistics based on the number of iteration for finding the global optimal solution.

The number of population $m_p$, the number of frogs in the population $n_p$ and the number of local iterations $N_f$ all have significant effects on the computation time of the algorithm and all increase as the parameter values become larger. However, the number ($q$) of frogs in the subpopulations constructed in each population had no significant effect on the calculation time. All the four parameters have influence on the success rate of the algorithm, the most important one is $m_p$. If the value is too small, the success rate will be lower, and the success rate will be stable at 100% with the increase of its value. The local iteration number $M$ also has a great influence on the success rate of the algorithm, and it is not suitable to be too large or too small. In the process of numerical change of $n_p$, the selected parameters, it has little influence on the success rate of the algorithm, which is more than 80%. As the value of $q$ increases, its success first rises and then falls.

5 CONCLUSIONS

(1) For the lack of uniform theoretical model in optimal sensor placement in diagnosis system, this study proposes a mathematical model for optimal sensor placement of gearbox fault diagnosis system based on the causality between sensor and system fault mode with the system test performances (fault detection rate, resolution rate, isolation rate, and virtual universal rate) as constraints, to realize the optimal selection of the sensor network location.

(2) For 0-1 integer optimization, the crossover operator and mutation operator, individual updating strategies, in genetic algorithm are introduced into the updating strategy of SLFA to form a new discrete SFLA, and the global convergence of the improved algorithm is theoretically demonstrated according to the properties of the homogeneous Markov chain.

(3) With excellent global optimization ability and fast sensitivity, this algorithm can effectively solve the optimal placement model of gearbox sensors proposed in this study. At the same time, it also provides an intelligent method for solving other related 0-1 type NP difficulties.

6 REFERENCES


