Semantic Artificial Immune Model for Fault Diagnosis

Chu-jiao Wang, Shi-xiong Xia, Yong Zhou
School of Computer Science and Technology
China University of Mining and Technology
Xu Zhou, China
wcj@cumt.edu.cn

***Abstract*—Applying artificial immune system to fault diagnosis is a new development direction in artificial intelligence, but the traditional artificial immune mode could not reasonably reflect the semantic similarity in the complexity problem space. For issues of semantic description of fault diagnosis, this paper introduces group cooperative mechanism of lymphocyte with a semantic tag to artificial immune system, thus solves the problem of semantic logical reasoning of fault knowledge. This paper presents a semantic-based artificial immune diagnosis model; designs an immune negative selection diagnostic in semantic environment; utilizes new coevolutionary algorithm diagnostic; and diagnoses fault in large electromechanical devices. The experimental results show that the method used in this paper has higher classification accuracy compared with the traditional artificial immune diagnostics, at the same time, verified expression capacity of semantic-based lymphocytes, which can provide more valuable diagnostic information.**

***Index Terms* —Fault diagnosis; Artificial immune system; Lymphocyte; Semantics**

I. INTRODUCTION

Usually a premise of solving the problem of fault diagnosis is a single fault assumption, which, therefore, can lead to a decline in diagnostic efficiency and accuracy. A basis of general fault diagnosis is building mathematical models, while under actual operating conditions, it is very difficult to obtain a precise mathematical model and that also requires strong expertise and domain knowledge [1]. However, as diagnosis method based on expert knowledge and data mining requires a lot of empirical data on the one hand, due to the representation of knowledge is usually based on binary logic with Boolean operators or decimal system, which is not a true reflection of many complex problems in the real world on the other hand. The knowledge-based fault diagnosis system mainly includes the fault diagnosis method based on deep knowledge and one based on shallow knowledge, including: rule-based fault diagnosis, fault diagnosis method based on fault tree, fault diagnosis method based on fuzzy set, the fault diagnosis method based on artificial neural network, fault diagnosis method based on support vector machine, and fault diagnosis method based on rough set [2-9].

Artificial immune technology has obtained considerable development in the last century and early this century, and a great progress in the area of artificial immune technology has brought new opportunities and new ideas for solving complex system fault diagnosis problem. Currently, prestigious universities abroad have done some effective studies in the fault diagnosis problem using artificial immune method [10-15], and Chinese scholars also carried out some works in this area [16-19].

A. The Problems Faced with Fault Diagnosis

The existing general method of model abstraction from monitoring data is first convert the data into some kind of advanced data representation, such as the symbol sequence or a point in feature space, and then do clustering (or classification) in the symbol sequence or feature space to generate the mode or mode collection. The key issue of problem is the data representation.

In situation of mass run data on large and complex devices, it will be inefficient that using the original data directly to process content-based query and clustering and classification analyses, and ultimately to realize the fault mode identification. The difficulty to solve problem is that the information content has no better formal representation; the data are lack of clear semantic information; and data format possess heterogeneity and semantic multiplicity, so that computer cannot intelligently understand the specific meaning of data, and therefore computer automatic processing cannot be achieved. Furthermore, data are isolated each other or lack of fine-grained semantic association.

Therefore, it is needed to study the appropriate data representation to process data on a higher level of data representation. This kind of treatment not only achieves data compression and reduces the computational cost of the subsequent processing, at the same time, but also describes the raw data in a more abstract level, thus is conducive to discover laws. The information description based on semantic association becomes an effective method for solving this problem. To compensate for the current lack of computers in data understanding, it is a

doi:10.4304/jcp.8.8.2059-2068

good idea to give monitoring data explicit semantic information, and thus to establish the temporal knowledge representation fault model based on semantic association. Therefore, for the problem that existing fault diagnosis system cannot take full advantage of a potential value of semantic information in monitoring data, this paper introduced the problem of the semantic representation of the fault modes, and proposed knowledge base based on the semantic description to represent fault information, in which the purpose is to more effectively extract useful fault knowledge from large fault information resource, and to achieve a more effective diagnostic reasoning.

B. *The Existing Problem of knowledge Representation in Traditional Artificial Immune Model*

The characteristic identification of the immune system, self-organization, memory, adaptation, and learning ability are needed for handling complex equipment fault diagnosis problem. In the research process of applying artificial immunology to fault diagnosis, scholars have proposed a lot of artificial immune models [20-24], including reverse selection model, clonal selection model, chromosome model, and immune network model, but there exist some common problems in these models needed to be resolved, that is, the traditional artificial immune model exists some problems on knowledge representation.

Artificial immune system based on the binary represenation cannot reflect the semantics of complexity problem space, and its efficiency is not satisfactory in the use of reverse selection algorithm covering nonself space. Therefore, in certain applications, it is needed to train the system to generate large number of diagnostics to strengthen coverage effects. In addition, the knowledge based on the binary representation makes the distinction between normal and abnormal just too clear, and is unable to meet the requirements of certain areas of fuzzy interval division [25,26].

Applying multidimensional real number to its antigens solves the defect, in a certain degree, that binary model cannot achieve fuzzification. Compared to the binary representation method, the real number representation method increases the degree of knowledge representation, and therefore, also increases the possibility that diagnostic extracts knowledge of the stronger representation ability. But because the search space is often continuous in the real number representation, it is impossible to use enumeration combination to achieve mathematical analysis, so it is relatively lack of inquiry ability. And when using negative selection principle to achieve fault diagnosis, like binary encoding, the real-coded also exists the problem that generated diagnostic has no enough coverage for testing space.

It is necessary to construct a more effective antigen and lymphocyte coding model and matching algorithms, which can better reflect the semantic similarity in complex problem space. Based on the above analysis, we propose a "semantic-centered" artificial immune fault diagnosis method. Through established fault ontology model, the semantic editing method of artificial immune system is proposed to reflect the important logic relations of fault knowledge, and to describe the characterization of the system behavior under normal and abnormal patterns. On this basis it is possible to discuss artificial immune fault reasoning methods in semantic coding model, and to focus on solving the reasoning model for negative selection algorithm, improved algorithm, and fault diagnosis in semantic space. At the end, the paper tested the robustness and effectiveness of the system. The semantic-centered artificial immune fault diagnosis method integrates both technologies of immune algorithm and semantic knowledge, which is a promising task consistent with the nature of things.

II. Group Cooperative Artificial Immune Model Based on Semantic Representation

A. *Biological Lymphocyte Costimulatory Principle*

The biological immune costimulatory principle provides a source of innovation for the artificial immunology to solve complex system fault diagnosis problem. From viewpoint of the organizational structure, the biological immune system is a multi-layered defense system. Artificial immune system is proposed based on the adaptive immune mechanisms, and lymphocytes were divided into two types of cells, in which there is complementary mutual stimulation between B cells and T cells.

Introducing biological costimulatory characteristics into fault diagnosis provides a new way of thinking for the traditional fault diagnosis methods:

I. Pre-classifying the abnormal perception information, and extracting antigen multi-dimensional attribute representation to solve the problem of the uncertainty of the diagnostic object;

II. Achieving group lymphocyte model definition, structure, and evaluation to solve the problem of diagnostic method uncertainty.

Most of the fault diagnosis systems achieve fault diagnosis based on the fault symptoms of fault mode, in which diagnostic technology is relatively mature, and the known fault modes can be accurately diagnosed. However, its main drawback is difficult to find the unknown new fault modes and multiple fault modes. The biological lymphocyte costimulatory principle uses a combination of innate immunity and adaptive immunity, and uses innate danger signals to induce acquired immune response, which possesses ability to diagnose unknown antigens. Thus, learning from the principle in the process of fault diagnosis can eliminate the diagnosis uncertainty, and especially for unknown fault mode, it can improve diagnostic accuracy.

B. Description of Fault Diagnosis Problem in Semantic Space

In order to have a unified description for artificial immune diagnosis model in semantic space, here gives a specific definition.

Definition 1: A system is defined as the ordered pair $(AttriSet(Rel^n), Rel^n)$, and $AttriSet(Rel^n)$ represents logic description of all symptoms of components on various levels in the system.

where Rel^n is an n-dimensional data set, n≥1. The characteristic pattern of Rel^n in semantic representation model is $Rel^n(A_1, A_2, \cdots, A_n)$. Denote $AttriSet(Rel^n) = \{A_1, A_2, \cdots, A_n\}$ as attribute set of Rel^n, and denote $Dom(A_j)$ as domain of attribute A_j, where $j = 1, 2, \cdots, n$. Tuple $t = (V_{A_1}, V_{A_2}, \cdots, V_{A_n}) \in Rel^n$, where $A_j \in AttriSet(Rel^n)$ and $V_{A_j} \in Dom(A_j), j = 1, 2, \cdots, n$.

Definition 2: Fault location inclusion relations can be expressed by <SP; SP means fault components contained in model; and A|SP| denotes description of an abnormal component. Given model A1 and A2, if A1|SP| ⊂ A2|SP|,then A1<SP A2, where A1 and A2 are the model of A.

Definition 3: If D is defined as a minimal diagnosis, then D = (A|SP|, Rel^n ,FS), where FS represents an observation of system state and behavior.

Definition 4: Complex system diagnosis problems can be described by a quintuple <D,A,C,M,FS>.

where $C \subseteq D \times A = \{C_1, C_2, \cdots, C_L\}$ represents a causal mapping relationship collection from D to A, and $M \subseteq A \times D = \{M_1, M_2, \cdots, M_L\}$ represents a causal mapping relationship collection from A to D. $\{Ci =< d_k^{\ i}, a_j^{\ i} >| d_k^{\ i} \in D_i, a_j^{\ i} \in A_j, i = 1, 2, \ldots, L\}$ d enotes a casual mapping relationship from fault to symptom at all levels, and $< d_k^{\ i}, a_j^{\ i} >$ represents that the k-th fault $d_k^{\ i}$ occurrence in i-th level component may cause the j-th symptom $a_j^{\ i}$ of the same level. $M = \left\{< a_l^{i+1}, d_k^i >| d_k^i \in D_k, a_l^{i+1} \in A_l, i = 1, 2, \ldots L\right\}$ represents casual mapping relationships between a collection A_{i+1} of lower layer component fault symptoms and a collection D_i of upper layer component fault symptoms, and $< a_l^{\ i+1}, d_k^{\ i} >$ indicates that the first symptom a_l^{i+1} in i+1-th level component may be the fault cause for k-th fault $d_k^{\ i}$ occurrence in i-th level component.

D_i and A_i denote i-level fault set and fault symptom set; L represents the level of depth of complex systems; relation C characterizes the causal mapping relations from fault to fault symptoms at all levels.

For any $a_j^{\ i} \in A_i$,its fault mapping is $D(a_j^{\ i+1}) = \{d_k^{\ i} |< d_k^{\ i}, a_j^{\ i} >\in C_j\}$. For any $a_j^{\ i+1} \in A_{i+1}$, its fault mapping is $D(a_j^{\ i+!}) = \{d_k^{\ i} |< a_k^{\ i+1}, d_k^{\ i} >\in M_i\}$. Similarly, the mapping between the fault symptoms and fault causes can be defined as

$\{A(d_k^{\ i}) = \{a_l^{\ i} |< d_k^{\ i}, a_l^{\ i} >\in C_i\}$,

$\{A(d_k^{\ i+1}) = \{a_l^{\ i+1} |< a_l^{\ i+1}, d_k^{\ i} >\in M_i\}$.

In order to achieve an accurate diagnosis for system fault, the first thing to do is to make an effective description of system behavior. Literature [25] established the spacecraft system model to determine the possible fault using the first-order logic diagnosis. In the literature [26], the MPL model editing method is similar to the Clips syntax. It becomes a key of the semantic-based artificial immune fault diagnosis method that how to take advantage of the existing fault ontology knowledge model to achieve artificial immune system editing program, establish the semantic logic description method for the system, and try to make use of existing mature tools for solving problem.

Definition 5: Fault rules can be expressed with well-formed formula composed of well-defined atomic predicate formula. Attribute sets are $X \subset AttriSet(Rel^n)$ and $Y \subset AttriSet(Rel^n)$, $X \neq \varnothing, Y \neq \varnothing, X \cap Y \neq \varnothing$. Let P and Q are composed with X and Y attribute symbols and logical operations $\wedge, \vee, \neg$, then 1) the expression $R:"P \rightarrow Q"$ is called Rel^n meta-rules (MR), abbreviated to R_m, and denote $\{R_m\}$ as a collection of meta-rules; 2) P is called the rule antecedent and Q is called the rule after pieces; 3) P and Q are called R_m constitution primitive, and {P, Q} is R_m base element set; 4) Meta-rules constituted by the same primitive set is called homologous; 5) $W = X \cup Y$ is the meta-rules attribute set of R_m on Rel^n; 6) if $r_m, r_m' \in \{R_m\}$, and their meta-rule attribute sets are the same, then r_m and r_m' are called congeners. Compared with traditional meta-rules, definition 5 added in "or" and "not" logical operator, which makes meta-rule more expressive.

Assuming a meta-rule $R_m:"P \rightarrow Q"$, assign corresponding tuple t property values to all semantic properties in its expression in accordance with the corresponding property name, then constitute expression $"P_i \rightarrow Q_i"$ denoted as R_i, then 1) call R_i as meta-rule

instance of R_m on Rel^n, referred to as an instance; 2) assume attribute set W be a set of attributes in the rule antecedent and after piece expression, according to definition 5, known $W \subseteq AttriSet(Rel^n)$, there is a list $S = \langle a_1, a_2, \cdots, a_m \rangle$, in which $a_i \in W, i = 1, 2, \cdots, m, m = |W|$, denote S as a meta-rule attribute list composed by W; 3) call tuple $t_f = \prod_s \left(V_{A_1}, V_{A_2}, \cdots, V_{A_n} \right)$ as a characteristic tuple of instance R_i, for R_m on Rel^n, t_f can only uniquely determine one instance; and 4) refer $Instance(R_m, Rel^n)$ as an instance collection obtained after R_m on Rel^n instantiated.

III. Evolutionary Algorithm of Semantic-Based Artificial Immune Model

Why aplying AIS to fault diagnosis is because there is a lot similarity between anomaly detection mechanism of fault diagnosis and the diversity and dynamic coverage of immune system, which can well solve the dynamic and complex problem that general method cannot deal with.

The essence of the immune system is independent abnormal vivo monitoring system. In the immune system, there are mainly B and T two types of lymphocytes, in which T-cell maturation needs to undergo a negative selection process. The recognition of immune system themselves-nonself refers to that after several negative selection T cells are able to accurately identify non-hexyl, which is a process of learning themselves to generate a memory for non-hexyl. Therefore, the most important feature of the immune system for fault diagnosis is using a limited number of diagnostic devices to identify an unlimited number of faults.

A. Deficiency of Traditional Negative Selection Algorithm in Fault Diagnosis Application

Forrest and her team studied a negative selection algorithm of change detection in 1994. Generally speaking, the negative selection method is using defining self-space to train diagnosis device, and then using a dynamic update to obtain diagnosis device that can detect the unique faults. Forrest's original algorithm used equal length binary bit string to represent behavior patterns, and used domain method to make affinity match. Since Forrest proposed negative selection algorithm, it has been a very good application in fault diagnosis. The advantage of negative selection algorithm of binary encoding applied to fault diagnosis is that a limited number of diagnostic devices can be used to judge unlimited type of faults, and this implementation method is particularly suitable to the system that can successfully convert a problem into discrete binary.

However, negative selection algorithm in some areas exists serious scalability problems, especially when directly is applied to fault diagnosis, in which the root of the problem lies in the knowledge manifestation and matching rules of the traditional negative selection algorithm makes the problem solving only act on the genotype level. The representation of real value is superior to the binary representation of a binary, which solves defects of representation of a binary in some respects. However, the real number coding has some fundamental issues, such as not reflecting semantic association of problem space, a large number of redundant of diagnosis device, and less capable of analyzing deep-seated fault semantic reasons.

B. Improved Semantic-Based Negative Selection Algorithm

Expressing knowledge by ontology is favorable for reuse and reasoning of knowledge. The structure of domain ontology for Negative Selection Algorithm was formally defined, and the relations between ontology elements and semantic analysis were illustrated[27,28,29]. The system status eigenvectors obtained by the monitoring objects under normal operating conditions constitute a collection of self-mode, denoted as Vs, $Vs = (Vs_1, Vs_2, \cdots, Vs_n), 0 \le Vs_i \le 1, i = 1, 2, \cdots, n$. The system status eigenvectors constitute obtained by the monitoring objects under abnormal operating conditions constitute a set of non-self mode, denoted as $V_N = (V_{N1}, V_{N2}, \cdots, V_{Nn}), 0 \le Vs_i \le 1, i = 1, 2, \cdots, n$. In M-dimensional space, the lymphocyte collection constituted by M N-dimensional state-detection normalized vectors (lymphocytes) $B = (ab_1, ab_2, \cdots, ab_n), 0 \le b_i \le n, i = 1, 2, \cdots, n$ is denoted as $Dia = \{B_1, B_2, \cdots, B_n\}, j = 1, 2, \cdots, n$, which is called state detector, where ab_i represent lymphocytes.

In order to maintain the diversity and synergy of multi-lymphocyte model, on the basis of analyzing limitations of the negative selection algorithm, this paper learned from the idea of the traditional negative selection algorithm, and proposed a new improved negative selection algorithm suitable for facility fault diagnosis, and its main idea is:

(1) Diagnose whether the system appears anomaly by constituting self-collection;
(2) Diagnosis devices with different fault modes only match non-self collection with different modes, which can diagnose different types of faults, with a relatively small amount of diagnosis devices for achieving the purpose of unlimited types of fault diagnoses;
(3) Propose a negative selection algorithm expressing semantic space, in which the matching algorithm of its corresponding semantic representation reflects more intuitive association of individuals in the problem space.
(4) Use immune particle swarm algorithm in the training model to train each lymphocyte

model;

(5) Train each generation of the algorithm to generate multiple collections of recognizers, which is produced by lymphocytes model;

The specific algorithm is as follows:

(1) Describe autologous information and autologous string collection is established during the normal system operation time. Define autologous ss a length L of the collection of strings S in length L, the expression is F_1 Ac_1: $Fs_1 \wedge Fs_2 \wedge \cdots \wedge \neg Fs_{k+1} \cdots \wedge \neg Fs_m$;

(2) Generate the diagnostics collection under different fault modes. The system randomly generates a length L string as the initial diagnostic device, and screen diagnosis collection in accordance with matching algorithm in the form of semantic space.

Step 1. Set the initial parameters of the algorithm; randomly generate initial lymphocyte population A(0); let current iteration k: = 0;

Step 2. Perform a cloning operation on lymphocyte population A(k); obtain a new lymphocyte population $A^{(1)}(k)$;

Step 3. Perform immune genetic manipulation on lymphocyte population $A^{(1)}(k)$; obtain a new lymphocyte population $A^{(2)}(k)$;

Step 4. Calculate the affinity between all lymphocytes of the semantic representation;

Step 5. Perform clonal selection operation on the lymphocyte population $A^{(2)}(k)$ and A(k), obtain a new lymphocyte population;

Step 6. Let k: = k +1; if satisfy termination condition, the algorithm stops; otherwise, return to Step 2;

(3) Monitor system with the collection of diagnostic device. If test samples are found matching with some kind of fault pattern, the system may have appeared fault behavior in some kind of modes, and need further judgment and processing.

The important feature of the improved semantic-based negative selection algorithm is that the system behavior patterns use coding method in the physical sense and more knowledge skill semantic space. This paper will directly apply the concept of fault ontology to negative selection algorithm to construct diagnostic device, in which the matching rules are directly comparing individual semantic similarity in the problem space, which conforms semantic computation of stable model, and solves the scalability problem of the negative selection algorithm [17].

Immune response process of the system over external antigen is also a dynamic evolution and learning process of lymphocyte collection in system, wherein the matches between antigen and lymphocytes or between lymphocytes and lymphocytes perform based on the degree of the affinity between them, and immune molecule affinity is defined as the degree of similarity between the immune molecules.

Its unique strategy of improved negative selection algorithm under the semantic space is to treat the generated lymphocyte cells as autologous cells. This way can avoid repeatedly generating equivalent fault diagnosis devices, and can introduce certain filtering method to avoid invalid diagnosis devices. In specific applications, and similarly negation rule gene pool can be filtered to avoid generating semantic conflict or not sensitive rules.

C. Semantic-Based Affinity Assessment Method

Under normal circumstances, the affinity between antigen and lymphocytes or between lymphocytes and lymphocytes is calculated using the Hamming distance or Euclidean distance between the two. In the case of binary encoding Hamming distance is used and Euclidean distance is used in the case of real-valued coding.

a) General Method of Immune Match

In immune algorithm, matching rule between antigen-lymphocyte cells directly guide the evolution of the immune system and ultimately immune recognition and judgment, therefore, matching rules is a key point in the normal functioning of the immune system. On training diagnostic phase, matching rules is used to judge diagnosis rationality and distribution, while on fault diagnosis stage, matching rules is directly used to judge whether the fault has happened and what type of fault has happened.

In the immune system, matching can be classified exact match and partial match. Typical partial match rules are: the Hamming rules and r vicinal match rules.

b) General Method of Semantic Match

Noumenon as a tool of semantic description and building conceptual model makes the matching algorithm evolve from formal matching process to the content and semantic matching, which overcomes the many defects of simple binary or decimal coded representation. There are two categories of semantic matching methods proposed from different angles from existing research results: the semantic matching model based on the information

content and semantic matching model based on geometric distance.

Specifically Semantic Match algorithms mainly have Overlap methods and VDM (Value Difference Metric) method, as well as algorithms evolved on the basis of these two algorithms [17].

c) Semantic-Based Immune Affinity Degree Matching Algorithm

In order to effectively combine immune system and ontology semantic, this paper uses the domain ontology semantic similarity to measure the degree of attributes match of the immune system, on the basis of the analysis of immune matching mechanism in the immune system. In the immune affinity matching algorithm design, taking into account the semantic encoding of immune cells, the continuous r bits matching rule is used to achieve immune matching mechanism, meanwhile the ontology semantic matching model is used in immune matching mechanism, and ontology semantic similarity measure method is applied to immune matching model.

At the same time in achieving the semantic-based immune matching algorithm, the design approach of parallel computing is used to improve the capacity and efficiency of the operation of the system, and this in itself is also in line with the parallel characteristics of the natural immune system.

The definition of continuous r bit immune matching rules based on semantic similarity is: if two vectors to be matched each other, there must be at least r continuous attribute semantic similarities on corresponding bit matching. Therefore, according to the previously defined antigen lymphocyte representation, we need to determine whether Ag.Fs, and Ab.Fs match each other, that is, whether match(Ag.Fs Ab.Fs). Thus, it is possible to calculate separately match(Ag.Fsi,Ab.Fsi) for each attribute to deal with semantic attributes.

Define the degree of match between the lymphocytes Ab_i and the signs to be tested Ag_j, that is, the affinity as:

$$Aff\left(Ab,Ag\right)=\frac{1}{\sqrt{L}}\sqrt{\sum_{i=1}^{L}match\left(Ab_i,Ag_j\right)^2}$$

$$match\left(Ab_i,Ag_j\right)=match\left(Ab.Fs_i,Ag.Fs_j\right)$$

$$=\begin{cases}0,\text{attribute i, j not match}\\1,\text{ attribute i, j match}\end{cases}$$

where, Aff is in range between 1 and 0, and the bigger the Aff is, the better match Ab_i and Ag_j are.

During a specific process of lymphocyte evolution, the degree of similarity between the lymphocytes is needed to determine the promotion and inhibition process of lymphocytes, and the degree of similarity between lymphocytes characterizes the phenomena of collaboration and exclusion between the different diagnostic devices. Suppose two lymphocytes Ab_i and Ab_j represent diagnostic devices of two different fault modes, then the definition of their degree of similarity S_{ij} is:

$$Sim\left(Ab,Ag\right)=\frac{1}{\sqrt{L}}\sqrt{\sum_{i=1}^{L}match\left(Ab_i,Ab_j\right)^2}$$

$$match\left(Ab_i,Ag_j\right)=match\left(Ab.Fs_i,Ab\sum_j .Fs_j\right)$$

$$=\begin{cases}0,\text{attribute i, j not match}\\1,\text{ attribute i, j match}\end{cases}$$

S_{ij} is in the range between 0 and 1, the greater the S_{ij} is, the greater the similarity is between Ab_i and Ab_j, which means the more obvious on the degree of inhibition; on thr other hand, the smaller the S_{ij} is, the smaller the degree of similarity between Ab_i and Ab_j, which means the stronger role in promoting.

IV. Fault Diagnosis Model Based on Immune Response Mechanism

In a specific complex system fault diagnosis, monitoring equipment status signs are corresponding to antigens, while the fault detection collection is corresponding to lymphocytes. Self set S can be obtained by training the antigen set in normal state, and, similarly, the corresponding fault detection collection can be obtained by training the antigen set in fault mode, i.e. memory lymphocytes set NS, where, the all lymphocytes are able to identify the structure of such a fault antigen. The members in memory lymphocytes set are pairwise mutually disjoint, in which all memory lymphocyte sets are composed of memory lymphocytes library Lib, IM represents complete works, i.e.: $N_{si}\cap N_{sj}=\varnothing, Lib=N_{s1}\cup N_{s2}\cdots N_{sn}$, where i, j denote different fault modes, respectively. When the fault diagnosis system is corresponding to the entire immune system, there exists $S\cup Lib\subset IM, S\cap Lib=\varnothing$.

Figure1 shows the main structure of the system.

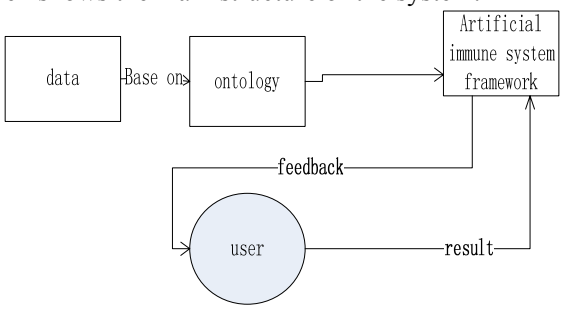

Figure 1. Diagram of Semantic-Based Artificial Immune Fault Diagnosis System

Due to the occurrence of the fault is usually related to many aspects of the machine, such as the mechanism of hoist itself and the production environment and the maintenance. If only one subsystem is considered, which will lead to incomplete information, and affect the final diagnostic results, and therefore, these subsystems should be combined to fully and accurately diagnose fault. Based on the ontology knowledge in area of established fault diagnosis system, the information integration between different systems is considered, so that the information can be shared between systems, as shown in Figure 2.

The system first constructs gene libraries under different operating modes from the training data of the fault mode and normal mode, and generates the initial lymphocytes from the gene libraries, which avoid producing a large number of lymphocytes without actual semantic meaning. First a predetermined number of the initial populations are generated in the training process. When the initial population is generated, it is needed to test the individual to determines whether it matches self=collection and the individual matching self-collection needs to be deleted until continuing to generate a new individual. Then the dynamic clonal selection is carried out on initial population to generate mature fault diagnosis, and diagnosis requires to cover nonself collection as much as possible. After completion of the training phase, fault diagnosis can be put in real-time diagnostics. If the fault cannot be diagnosed after a long time, the gene libraries need to have variation generalization. The following gives the framework of semantic-based artificial immune fault diagnosis system as shown in Figure 3.

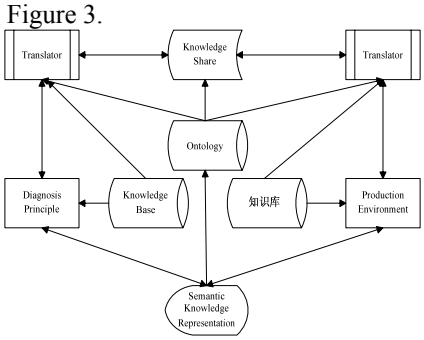

Figure 2. Semantic-Based System Diagnosis Model

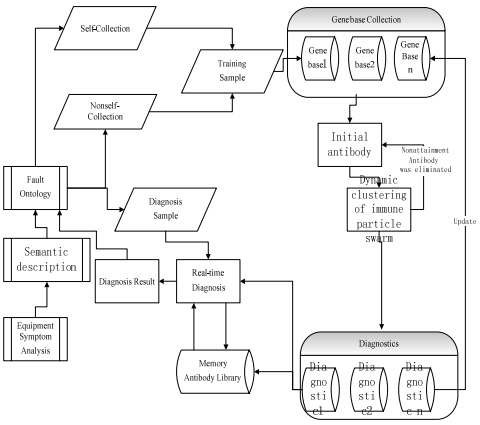

Figure 3. Semantic-based Artificial Immune Fault Diagnosis Model

V. EXPERIMENT DATA AND ANALYSIS

A. *Experiment Data*

According to different requirements for system diagnostic capability, produced danger signals may be from system level, subsystem level, component level, element level. The features can be extracted from various types of fault samples, especially those samples with characteristics of "danger", from the level of the subsystems, components, and element level.

The diagnostic reasoning experiment of this research topic takes Matlab6.5 as platform to find information from hoist fault ontology knowledge base and to achieve encoding conversion from antigen to lymphocyte. We selected motor fault examples in literature to validate the feasibility of our approach, and three common fault examination methods are selected to verify the ability to identify known type of fault and the unknown fault. The three kinds of typical faults are: initial mass imbalance of rotor, rotor bending, and rotor grinding and contacting. When different fault types occur, eight empirical cases corresponding to fault types, respectively, are shown in Table 1, and each row in table corresponds to a fault case, in which the element values in a row represent different fault symptoms.

This paper takes the established fault ontology of a hoist motor as an instance to achieve semantic conversion from antigens to lymphocytes, focusing on research and verification from the point of view of the diagnostic methods. Three kinds of fault modes of eight empirical cases mentioned above are selected. Table 2 lists all fault symptoms corresponding to the semantic representation of empirical cases.

TABLE I. .

EMPIRICAL CASES

Fault Category	Empirical Case
Initial mass imbalance of rotor	(2,6,11,18)(6,11,18)(11,14,18,25)(3,11,18) (2,11,14,1)(11,14,18,25)(6,11,14,18),(2.6.11)
Rotor Bending	(12,15,23,32)(12,19,23,32)(6,12,23,32)(3,15,19,25) (6,15,23)(12,19,23)(19,25,32)(3,12,23)
ZWZ Wear	(3,10,16,31,33)(10,16,20,22)(10,16,20)(3,10,31,33) (10,20,22)(10,16,20)(10,16,20,22)(10,15,20)

TABLE II.

FAULT SYMPTOMS

No.	Fault Phenomenon
2	Axis orbit is elliptical
3	Axis orbit is banana shape or a long strip
5	Spectrum has baseband
6	Spectrum has clear baseband
10	Spectrum has spread spectrum from low frequency to high frequency
11	Spectrum has no 2X and 3X

Therefore, describing some common fault events of the mine hoist based on fault ontology model can easily link fault event description and a description of the fault symptom, as shown in Table 3.

According to fault ontology and lymphocytes encoding method defined by this paper, the data in table 3 are converted to the semantic representation of the four-tuple, as shown in Table 4.

TABLE III.

FAULT SYMPTOMS OF CONVERTED FOUR-TUPLE FORM

No.	Fault Symptom	No.	Fault Symptom
1	Axes: contrail = oval	10	G-Frequency: change of component = little
2	Axes: contrail = banana-shaped, strip	11	G-Frequency: change of component =
3	Spectrum:Fundamental frequency component = exist	12	G-Frequency: change of component =
4	Spectrum: spread spectrum from low frequency to high frequency =	13	B-Frequency: change of component =
5	Spectrum: 2X and 3X component = little	14	B-Frequency: change of component =
6	Spectrum: mail, 2X and 3X component = exist	15	B-Frequency and Speed: contrail =
7	Amplitude: change of amplitude = little	16	Rotor: differential expansion = Increase
8	Amplitude: change of amplitude = Increase	17	Rotor: off-center = Increase
9	Amplitude: change of amplitude= Increase and unstable	18	Cylinder body: temperature differences

TABLE IV.

FAULT SYMPTOMS OF SEMANTIC FOUR-TUPLES

No.	Fault Symptom	No.	Fault Symptom
2	Axes: contrail = oval	18	G-Frequency: change of
3	Axes:contrail= banana-	19	G-Frequency: change of
5	Spectrum: Fundamental	20	G-Frequency: change of
6	Spectrum: Fundamental	22	B-Frequency: change of
10	Spectrum: spectrum from	23	B-Frequency: change of
11	Spectrum: 2X and 3X	24	H-Frequency: change of
12	Spectrum: mail, 2X and 3X	25	B-Frequency and
14	Amplitude: change of	31	Rotor: differential
15	Amplitude: change of	32	Rotor: off-center =
16	Amplitude: change of	33	Cylinder body:

B. *Simulation Result*

After multiple comparison tests, the following parameters are used for obtaining experimental results: training evolution iteration number iter = 100, each prospective diagnostic collection size N = 80, immune suppression threshold mP = 0.96, and match threshold as 0.84. The degree of similarity between the values in the partial sample can be directly observed from the Table5 and Figure 4, as a reference to set the immune threshold.

For comparison, the decimal artificial immunization method is used for diagnosis for same data collection. In experiment, the ten-fold cross-validation is used to estimate the accuracy of the semantic immune negative selection diagnosis and decimal immune-negative selection diagnosis, that is, each time five experiments are conducted. In each experiment, the same random established cross-validation collection is used for both of semantic immune negative selection diagnostics and decimal immune negative selection diagnostics, and two classifiers run in turn in one cycle in order to achieve the results. that is, with each cycle, two classifiers use the same data to train and test collection. The experimental results are shown in Table 6.

TABLE V.

SAMPLE SIMILARITY

1	1.	1. 3	1. 7	0. 60	1.	1. 7	1. 3	0. 80
2	0.	1. 3	1. 7	0.	0.	1. 7	1. 3	0. 80
3	0. 30	0. 30	1. 8	0. 30	0. 30	1. 8	0. 30	1. 8
4	0. 60	0. 30	1. 4	1.	0. 60	1. 4	0. 30	0. 80
5	1.	. 30	1. 4	0. 60	1.	1. 4	0. 30	1. 4
6	0. 30	0. 30	1. 8	0. 30	0. 30	1. 8	0. 30	1. 8
7	0.	1. 3	2. 1	0.	0.	2. 1	1. 3	1. 8
8	0.	0. 30	1. 8	0.	0.	1. 8	. 30	1. 8
9	0.	0. 70	0. 30	0.	0.	0. 30	0. 70	0. 60
10	0.	0. 70	0. 30	0.	0.	0. 30	0. 70	0.

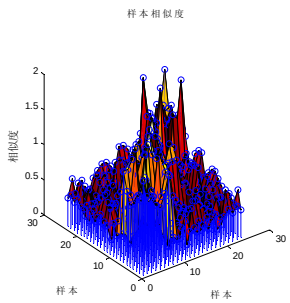

Figure 4. Similarity of samples

TABLE VI.

PERFORMANCE COMPARISON BETWEEN SEMANTIC IMMUNE NEGATIVE SELECTION DIAGNOSIS AND DECIMAL IMMUNE NEGATIVE SELECTION DIAGNOSIS

Diagnostic	Accuracy(%)	Variance	95%Confidence Interval	Error Rate(%)
Semantic Immune Negative Selection	76.5	0.027	0.051	0.025
Decimal Immune Negative Selection	70.3	0.045	0.080	0.091

The selected diagnostics based on the training results are the diagnostic for initial imbalanced mass of the rotor, the rotor bending diagnostic, and rotor grinding contacting diagnostic. As Table 7 shows that antigens 1-2, 3-4, and 5-6 belong to, respectively, faults of initial imbalanced mass rotor, rotor bending, and rotor grinding and contacting, which the correct rate of diagnosis is 100%.

TABLE VII.

DIAGNOSIS RESULTS OF TESTING ANTIGENS

No.	Matching	Threshold	[F1,F2,F3]
1	11,14,18,25	0.5	[0.71,0.01,0.01]
2	6,11,14,18	0.5	[0.75,0.02,0]
3	19,25,32,0	0.5	[0.04,0.62,0.]
4	3,12,23,0	0.5	[0.03,0.70,0.01]
5	10,16,20,22	0.5	[0.03,0.01,0.70]
6	10,15,20,0	0.5	[0.02,0.02,0.67]

According to the experimental results we can tell the known test samples of fault type, and the method proposed in this paper can determine directly its belonging degree to a known fault mode based on the number of activated diagnostics. When a sample of a new fault mode is generated, the present method does not require wide range of adjustments to collection of diagnostics, and only needs to train new sample to obtain diagnostics of the new fault mode. For the same situation, neural network learning methods must retrain the entire network, and conduct the large-scale adjustment of the original results. Meanwhile, as for decimal artificial immune diagnosis method, the description of the fault symptoms has more specific semantic meaning, so the finally derived fault classification has more specific judgments basis and reasoning path, and has more direct guidance to the final fault element-level positioning. Therefore, this method is more adapted to the large complex equipment fault diagnosis, and it is able to achieve the process of learning while diagnosing, which improves the reliability of diagnosis treatment results.

VI. SUMMARY

This paper discussed the logic semantic representation mode under immune system mode; analyzed defects of the traditional negative selection algorithm; proposed the negative selection algorithm based on semantic editing method; in order to express the rich semantic information of problem space, adapted the lymphocyte model; updated immune matching algorithm, and proposed semantic-based the r continuous bit immune matching algorithm.

The paper discussed in detail the semantic representation method of antigen and lymphocyte model in the immune system; used the semantic representation mode of the established fault ontology mode to propose effective immune cell genotype and manifestations; and gave encoding approach more suitable for the practical problems, thus improved the efficiency of generating lymphocytes.

REFERENCES

[1] QU Liang sheng,ZHANG Haijun. Some basic problems in machinery diagnostics[J]. China Mechanical Engineering,2000,11(2):211-216.

[2] ZHANG Cheng, LIAO Jian-Xin, ZHU Xiao-Min. Heuristic Fault Localization Algorithm Based on Bayesian Suspected Degree. *Journal of Software*,2010,21(10): 2610−2621(in Chinese with English abstract).

[3] ZHOU Fu na, WEN Cheng lin, CHEN Zhiguo. DCA Based Multi level Small Fault Diagnosis Method. Acta Electronica Sinica,, 2010,38(8):1874-1879.

[4] Purkait, P., Chakravorti, S.. Time and frequency domain analyses based expert system for impulse fault diagnosis in transformers. ,Dielectrics and Electrical Insulation, 2002,9(3): 433 – 445

[5] Li, B., Chow, M.-Y., Tipsuwan, Y.. Neural-network-based motor rolling bearing fault diagnosis . Industrial Electronics, 2000, 47(5): 1060 – 1069.

[6] Zhu Yongli,Limin, H., Lu Jinling. Bayesian networks-based approach for power systems fault diagnosis .Power Delivery, 2006,21(2): 634 - 639

[7] Wenli Yang,Lee, K.Y.,Junker, S.T.. Fuzzy Fault Diagnosis and Accommodation System for Hybrid Fuel-Cell/Gas-Turbine Power Plant. Energy Conversion, 2010,25(4): 1187 - 1194

[8] Rajasegarar, S., Leckie, C., Bezdek, J.C., Palaniswami, M.. Centered Hyperspherical and Hyperellipsoidal One-Class Support Vector Machines for Anomaly Detection in Sensor Networks . Information Forensics and Security, 2010,5(3):518 – 533.

[9] Dey, D., Chatterjee, B., ,Munshi, S.; Rough-granular approach for impulse fault classification of transformers using cross-wavelet transform . Dielectrics and Electrical Insulation,2008,15(5):1297–1304.

[10] J.H.Jung, J.J.Lee, and B.H.Kwon, Online diagnosis of induction motors using MCSA, IEEE Transaction on Industrial Electronics, 2006,53(6):1842-1852.

[11] de Lemos, R., Timmis, J., Forrest, S.. Immune-Inspired Adaptable Error Detection for Automated Teller Machines . Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2007,37(5):873–886.

[12] Carrano, E.G., Guimaraes, F.G., Takahashi, R.H.C.. Electric Distribution Network Expansion Under Load-Evolution Uncertainty Using an Immune System Inspired Algorithm .Power Systems,2007,22(2):851–861.

[13] Xin Wen, Brown, D.J., Qizheng Liao. Online motor fault diagnosis using hybrid intelligence techniques Industrial Electronics & Applications (ISIEA), 2010: 355 - 360

[14] Amaral, J.L.M., Amaral, J.F.M., Tanscheit, R.. ; An immune inspired fault diagnosis system for analog circuits using wavelet signatures Evolvable Hardware, 2004. Proceedings. 2004 : 138 – 141

[15] Aydin, I., Karakose, M., Akin, E.. Artificial immune inspired fault detection algorithm based on fuzzy clustering and genetic algorithm methods Computational Intelligence for Measurement Systems and Applications, 2008. CIMSA 2008. 2008 IEEE International Conference on 2008 , Page(s): 93–98.

[16] ZHENG Yong-huang,LI Ren-hou**.** Improved fault diagnosis algorithm based on immune network model. Control and Decision,2010,25(6):847-856.

[17] MA Zhan-fei; ZHENG Xue-feng. Concept and negotiation control model of Immune-SoftMan**.** Control and Decision, 2010,25(5):740-752.
[18] DU Haifeng ,WANG Sunan. Fault diagnosis of the reciprocating compressor based on ART-artificial immune network[J] . Chinese Journal of Mechanical Engineering,2002,38(4):88-90.
[19] DU Haifeng 1; 2 WANG Na 2 ZHANG Jinhua 2 SHAO Jie 3 WANG Sunan2 Fault Diagnosis Strategy Based on Complex Network Analysis**.** Chinese Journal of Mechanical Engineering,2010,46(3):90-96.
[20] D.Dasgupta, K.KrishnaKumar, D.Wong and M.Berry. Negative Selection Algorithm for Aircraft Fault Detection[C].Proceedings of the Third International Conference on Artificial Immune Systems (ICARIS-2004), Catania, Italy.September 13-16, 2004.
[21] Dasgupta D,Forrest S. Artificial immune systems in industrialapplications[C] // Proc.2nd International Conference on Intelligent Processing and Manufacturing of Materials,Honolulu,1999: 257-267.
[22] Nea,l M., Hunt, J., Timmis, J. Augmenting an artificial immune network [A]. Systems, Man, and Cybernetics, 1998.1998 IEEE InternationalConference[C]. 1998. 3821-3826.
[23] Hunt JE, CookeD E. LearningUsing an Artificial Immune Sys-tem[J]. Journal ofNetwork and ComputerApplications, 1996,19: 189-212.
[24] Timmis, J., Nea,l M. A resource limited artificial immune sys-tem fordata analysis[J]. Knowledge-Based Systems, 2001, 14:121-130.
[25] Brian CWilliansPPandurangNayak. A model based approachto reactive self-configuring systems[A].Workshop on Logic-Based Artificial Intelligence[C].Washington,June,1999.14-16.
[26] .Seung.HChung,Anthony Barrett.Distributed Real-time Model based Diagnosis[DB/OL].http://ase.jpl.nasa.gov/public/planning/papers/barrett-aero03.pdf,IEEE AC #1116,2003.
[27] T. Priebe, W. Dobmeier, N. Kamprath, "Supporting Attribute-based Access Control in Authorization and Authentication Infrastructures with Ontologies", Journal of Software, 2007, 2(1), pp.27-38.
[28] Peng Wu, SiKun Li. Social network analysis layout algorithm under ontology model. Journal of Software, 2011, 7(6): 1321-1328.
[29] J. Zhou, K. Watanuki. Skill ontology for mechanical design of learning contents. Journal of Software, 2012, 7(1): 61-67.

WANG Chu-Jiao was born in Macheng, Hubei, China, in 1976. She received master degree in control theory and control engineering from Xian jiaotong University, Xian, Shanxi, in 2004. Now she is the Ph.D. candidate of power electronics and electric drive department at China University of Mining and Technology, Xuzhou, Jiangsu, from 2006. Her research interests include data mining, artificial immunity algorithm, artificial intelligence and information security. She has published more than 10 papers in these areas. She is currently an university teacher in School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu, China. E-mail: wcj @cumt.edu.cn

XIA Shi-Xiong was born in Xuzhou, Jiangsu, China, in Sptember 1961. He received the Ph.D. degree in control theory and control engineering from China University of Mining and Technology, Xuzhou, Jiangsu. His research interests include data mining, genetic algorithm, artificial intelligence and wireless sensor network. He has published more than 30 papers in these areas. He is currently an Professor and tutor of postgraduate candidate in School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu, China. E-mail: xiasx@cumt.edu.cn