



# Journal of Applied Sciences

ISSN 1812-5654

**science**  
alert

**ANSI***net*  
an open access publisher  
<http://ansinet.com>

## **Sigmoid Model: A Simulation for Inference Process of Engineering Diagnosis**

Liangsheng Qu and Zhaoyong Hu  
Research Institute of Diagnostics and Cybernetics, Xian Jiaotong University,  
Xi'an, Peoples Republic of China, 710049

---

**Abstract:** The study emphasizes three kinds of diagnostic knowledge in engineering diagnosis and examines their structure according to their roles. Some requirements and possible measures to improve the diagnostic knowledge are suggested. Generally, information entropy is used as a measure of uncertainty of a system. The study provides broad diagnostic entropy and utilizes it to characterize the inference process of engineering diagnosis. Then sigmoid model and logarithmic model are used to simulate the process. Several simple examples show that the former is better than the latter. In order to verify it more, computer simulation and real diagnostic practice are discussed and they also revealed that the sigmoid model is more precise than logarithm model.

**Key words:** Engineering diagnosis, diagnostic knowledge, diagnostic entropy, sigmoid model, logarithmic model

---

### **INTRODUCTION**

Diagnostic practice shows that engineering diagnosis is crucial for continuous, high-effective, safe running and for foregone maintenance. It is especially important for large process equipment, which continuously operate without spare unit. Engineering diagnosis can decrease effectively appearance of great accident. The purpose of engineering diagnosis, in general, is to identify the potential faults inside an operating machine according to its physical symptoms and performance degradation, especially to recognize leading fault, which determines symptoms of the diagnostic object.

In fact, engineering diagnosis is an inference process, which diagnostician recognizes the inherent essence of diagnostic object based on his theoretic knowledge and practical experience about engineering diagnosis. Thus it is a process full of information and during it, diagnostic information is accumulated gradually till diagnostic conclusion is obtained.

As we all know, information entropy is a suitable quantitative measure of system uncertainty. The diagnostic entropy characterizes uncertainty of diagnostic object, which consists of those potential malfunctions. It is an intrinsic attribute of diagnostic object. That is to say, for a certain object, it is a constant quantity in some period and it does not vary during the inference process. To using diagnostic entropy to depict the process, we proposed concepts of broad diagnostic entropy, initial

diagnostic entropy and decreasing ratio of diagnostic entropy. The broad diagnostic entropy reflects experts' uncertainty about the possible leading fault inside the diagnosed object in his inference process. It will decrease gradually in the process and tend to a low level till definite diagnosis. At the same time, the initial diagnostic entropy shows initial uncertainty and the decreasing rate of diagnostic entropy characterizes the efficiency of diagnostic inference.

In order to describe the diagnostic process more distinctly, the study introduced sigmoid model and logarithmic model to depict diagnostic process through several simple examples. Computer simulation and real case study from an oil refinery are used to verify it, which revealed that sigmoid model is more efficient than logarithmic model.

### **DIAGNOSTIC KNOWLEDGE AND THEIR STRUCTURE**

There are three kinds of knowledge required when an expert performs a general problem-solving process. They are conceptual knowledge, problem-state knowledge and third, operational and procedural knowledge<sup>[1]</sup>. These three kinds of knowledge are strongly bi-directional connected to one another and form a complete knowledge structure. In diagnostic practice, we also need many kinds of diagnostic knowledge to determine the leading malfunction accurately. The symptoms of the diagnosed

---

**Corresponding Author:** Zhaoyong Hu, Research Institute of Diagnostics and Cybernetics, Xian Jiaotong University, Xi'an, Peoples Republic of China, 710049 Tel: +86 29 82669053 Fax: +86 29 83237910  
E-mail: zhyhu@mailst.xjtu.edu.cn

object, its past and present status and its design and manufacturing features, etc., all these consist of problem-state knowledge or diagnosed object-oriented knowledge (OOK). The 'if-then' rules of 'symptom-fault' type as well as some basic concepts and principles in physics and mechanics consist of conceptual knowledge (CK). Finally, different methodological knowledge (MK), for example, signal processing, fuzzy inference, computer simulation and field test, etc., consist of operational and procedural knowledge.

The conceptual knowledge of expert comes from his theoretical level and his long period of diagnostic experience. Since the engineering diagnosis belongs to reverse engineering, we have to identify the input (fault inside an operating object) according to the output (symptom) of partly black or gray box. The accumulation of this kind of conceptual knowledge is much more difficult. Secondly, the diagnostic problem is always a multi-solution problem. The same symptom may correspond to several kinds of faults and one fault may correspond to different symptoms. Thirdly, the expert always applies his conceptual knowledge implicitly and unconsciously in his diagnostic practice. Thus it is very difficult to gather his conceptual knowledge sometimes.

There are some suggestions to novices in order to widen their conceptual knowledge:

First, since we often need to evaluate the physical parameters quantitatively in diagnostic practice, their approximate normal ranges should be well recorded. Such parameters may be the critical speeds of certain type of machines, their normal bearing clearances, the approximate balancing weights applied on a centrifugal compressor, a large blower or a large steam turbine, etc.

Second, we are not only need to summarize the 'if-then' rules, but also should pay attention to the 'if not -then ', 'if-then not' or 'if not-then not' rules. Such rules may be in the form, for example, 'If the rotor vibration condition can't be improved after rotor being replaced, then probably all kinds of faults from rotor should be excluded without further consideration' or 'If the vibration amplitude changes insignificantly when the rotor passes through its critical speed, then the vibration probably does not due to its imbalance'.

These rules may become powerful evidence in diagnostic practice and could often be overlooked by novices.

Even though the experts possess relatively fruitful conceptual knowledge on equipment malfunctions and very rich methodological knowledge, but they still neither possess the knowledge on the operating conditions of certain diagnosed object, nor the knowledge about its historical records. In addition, the experts may not be

familiar with its original design considerations as well as its manufacturing process. All the knowledge listed above influence its final diagnostic uncertainty and should be acquired through the diagnostic process. For example, the historical record, in general, does not play direct role in inference process. However, it can be transformed into useful information according to the comprehension of experts.

So the diagnostic process isn't as simple as some people imagined that every potential fault in a running machine could be diagnosed by means of a simple vibration spectrum and all that we should do in engineering diagnosis is simply to transmit the sampled vibration signals. In diagnostic process, the expert needs to consummate his knowledge structure by adding Object-oriented Knowledge (OOK), process it by applying both conceptual and methodological knowledge (CK and MK) and draw diagnostic conclusions step by step through logical inference including deduction, induction, associate memory and other reasoning methods.

Above stated tells us, first, the expert's knowledge structure is not perfect and needs to be consummated in diagnostic process. Second, the main part of transferred information should consist of object-oriented knowledge, including:

The historical record or the case history of diagnosed object:

- The operation record after recent overhaul;
- The present operating parameters and physical parameters: e.g. vibration condition, thermal condition, machine sound, load, speed and bearing temperature, etc.
- Typical case studies of similar equipment;
- Design features, manufacturing and assembly process.

Third, engineering diagnostic system should be equipped with different toolbox, such as simulation toolbox, signal processing toolbox and should have the possibility for on-line experiments on test rigs and field tests of diagnosed object, such as observing the vibration level with the change of temperature of lubricating oil or input parameters.

Finally, a friendly discussion environment is advantageous for better inference. That is to say, engineering diagnosis should be performed under their cooperation between diagnostician and field technician.

In a word, the knowledge structure in engineering diagnosis is a type of 'OOK-CK-MK' structure with bi-directional links as shown in Fig. 1. The stronger the bi-directional links in OOK-CK-MK are, the more qualified the expert will be.

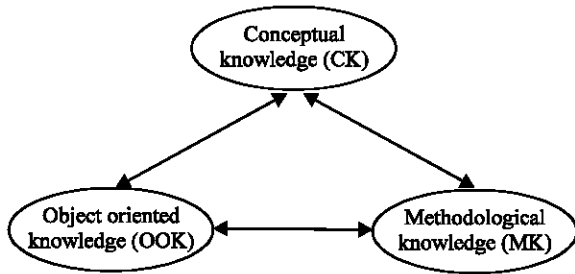


Fig. 1: The knowledge structure in engineering diagnosis

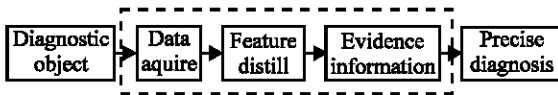


Fig. 2: A diagnostic process

The engineering diagnosis would be effective if all diagnostic information were useful knowledge. In practice, the diagnostic information is always noisy. It includes not only useful information, but also redundant and contradictory information. Let alone neither field practitioners, nor domain experts are always conscious of which information is useful in the beginning stages of engineering diagnosis. They become conscious only after they have been familiar with all the information they had received and begin to infer the root cause of certain symptoms according to the acquired knowledge.

### DIAGNOSTIC PROCESS AND DIAGNOSTIC ENTROPY

The engineering diagnosis can be parted into direct diagnosis and indirect diagnosis. The former means that it determines directly the status of some key parts, such as bearing clearance of principal axis. But direct diagnosis is always infeasible for the limitation of structure or circumstance. So indirect diagnosis is common and important, that is to say, we recognize the inner defects or the status of equipment based on those behaviors and symptoms of whole of the diagnosed object while not certain part of it. Thus the essential of engineering diagnosis is an inference process based on information. So we use entropy to characterize diagnostic process.

A diagnostic process can be depicted as shown in Fig. 2. From it, we could see that information is a main line throughout the whole diagnostic process. Especially, the parts in dashed pane may be performed many times till diagnostician can obtain precise conclusions. Thus, engineering diagnosis is a process of information accumulation and processing.

As we all know, information entropy is an important and common criterion characterizing the system

uncertainty. And reasonably, we can use it to describe process of engineering diagnosis, named diagnostic entropy<sup>[1]</sup>.

**Definition:** There are  $n$  potential faults in an equipment and accordingly, their probabilities are  $P_i$  ( $i=1, \dots, n$ ). Then diagnostic entropy is

$$H = - \sum_{i=1}^n P_i \log(P_i) \quad (1)$$

For equipment, the bigger diagnostic entropy is, the more uncertainty of the leading fault is, or the more potential faults are and then the more difficulty of diagnostic inference is. Thus, diagnostic entropy characterizes the diagnosed ability of equipment.

Naturally, diagnostic entropy is an essential property of equipment, so it does not vary during diagnostic process. Thus, in order to depict diagnostic process using entropy, we need to extend its signification and name it broad diagnostic entropy, i.e. diagnostic entropy in broad sense.

**Definition:** Diagnostician considers there are  $m$  potential faults in an equipment and their probabilities are  $P_i$  ( $i=1, \dots, n$ ). Then broad diagnostic entropy is:

$$H = - \sum_{i=1}^m P_i \log(P_i) \quad (2)$$

Here, from this definition, we can be aware that broad diagnostic entropy is defined from the point of view of a diagnostician. Thus, its value will be changed during the diagnostic process because diagnostician is aware of the equipment more and more clearly. So it can be used for diagnostic process. With the exception of especial illustration, the following diagnostic entropy refers to broad diagnostic entropy. Here, we need also give another two concepts: they are initial diagnostic entropy  $H_0$  and decreasing ratio of diagnostic entropy  $dH/dj$ . Initial diagnostic entropy consists of original or initial probabilities of those potential faults estimated by diagnostician without any evidence information. Decreasing ratio of diagnostic entropy means ratio of difference between the two adjacent diagnostic entropies during diagnostic process. They characterize a diagnostic process together. The former represents initial understanding of diagnostician about the diagnosed object, while the latter manifests efficiency of a diagnostic process.

For a same diagnostic problem, difference between an expert and a novice can be characterized using initial diagnostic entropy and decreasing ratio of diagnostic entropy. As an example, majority of grinding machines for

some time had been suffered from the malfunction in their hydraulic system. The expert knew from his past experience that the root cause of such malfunction was due to overheat of solenoid connected with the sliding valve and immediately suggested the practitioner to detect it. So the initial diagnostic entropy was low and the decreasing ratio of diagnostic entropy was very fast. On the contrary, the novice would diagnose the internal fault step by step with a high initial diagnostic entropy and low entropy drop rate. Another example is the illumination system of buildings. If the lamps suddenly put out, everybody would like to know if the fuse has been blown up. In this case, the system is an easily diagnosed design and the diagnostic entropy of inference process is extremely low.

In order to simplify the diagnostic process, we suppose that there exists only one leading fault inside the diagnostic object. Even though there may exist other potential faults inside the diagnostic object, the leading fault plays decisive role in performance degradation or poses a threat to its safe operation. Such a supposition is reasonable for majority of industrial equipment. We further suppose that the expert possesses fruitful theoretic knowledge, i.e., conceptual knowledge and methodological knowledge and plentiful practical experience on engineering diagnosis in advance, but lacks in object-oriented knowledge. So the latter should be obtained in diagnostic process.

Let us divide the inference process of engineering diagnosis into N stages. In each stage, new evidence is drawn, which may support a fault to be a leading fault or, may negate it. Furthermore, there also exists evidence, which can support or negate two or more faults simultaneously. Let  $\underline{E}_j$ -the evidence supporting j-th fault and  $\bar{E}_j$ -the evidence negating j-th fault, then an evidence stream will form in the process of engineering diagnosis:

$$\bar{E}_1 \underline{E}_2 \bar{E}_3 \dots \bar{E}_N \tag{3}$$

Since the obtained information stream comes in random order, it is reasonable to assume that each  $\underline{E}_j$  can locate in arbitrary order in the stream with equal probability and the whole stream composes of N stages.

Let us consider a simple case. Here, the expert lacks in any prior experience about the diagnostic object or similar equipment. According to Laplace principle, we should take equal probability as the prior distribution and the initial diagnostic entropy is equal to  $H_0 = \log(N)$ , where N is the number of possible faults. Suppose we have negated j-1 faults in j-th stage completely, then the diagnostic entropy should be decreased to  $H = \log(N-j+1)$ . Figure 3 shows the inference process in this simple case.

In each stage, after expert's decision, the probability distribution of all possible faults is redistributed: certain possible fault is negated and the probabilities of all remaining faults are increased and vice versa. The new distribution can be obtained according to minimum cross entropy principle<sup>[2,3]</sup>. Let  $p_i$ -new probability of I-th possible fault;  $q_i$ -original estimated probability of same fault according to prior information, we minimize the cross entropy or the Kullback-Leibler information number by using the Lagrangian operator L with a constraint:

$$\sum_{\text{all } p_i \neq 0} P_i = 1. \tag{4}$$

$$L = \sum_{p_i \neq 0} P_i \log \frac{P_i}{q_i} + (\lambda - 1) \left( \sum_{p_i \neq 0} p_i - 1 \right). \tag{5}$$

By putting  $\frac{\partial L}{\partial P_i} = 0$ , we have

$$\log \frac{P_i}{q_i} + \lambda = 0 \tag{6}$$

for all  $p \neq 0$ . From here we have:

$$P_i = \frac{q_i}{\sum q_i} \tag{7}$$

for all  $p_i \neq 0$ . If only one possible fault with original probability  $q_i$  remains under consideration in the last stage, then  $p_i = q_i / q_i = 1$ .

Generally, any evidence information cannot eliminate a fault absolutely. So we decrease its probability to a low level, such as 2%, if the evidence negates a fault. Thus the formula (7) will be adjusted as following:

$$P_i = \frac{q_i}{\sum q_i} \times \sum P_i' \tag{8}$$

$$\sum P_i' = 1 - \sum P(\bar{F}_j)$$

Where,  $\bar{F}_j$  represents those faults negated and  $\sum P_i'$  is summation of the probabilities of those negated faults.

Generally, the diagnostic process proceeds in an information system. With the accumulation of information obtained, the diagnostic entropy or the uncertainty is decreased from stage to stage. The total information gain in each stage can be expressed as the entropy drop from the initial stage to the current stage<sup>[4]</sup>.

### SIMPLE EXAMPLES AND COMPUTER SIMULATION

Figure 3 shows a simple example of diagnostic process. It is a very realistic example. So, generally, we

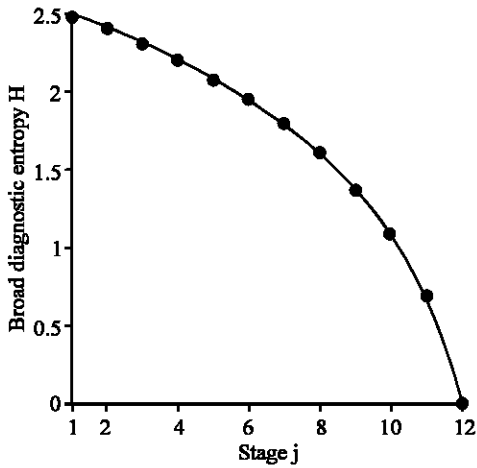


Fig. 3: A simple inference process of engineering diagnosis

change the simple example into another one. Here, we let one of faults keep a low level throughout the whole inference process of engineering diagnosis, such as 5% and all others have equiprobable distribution. At the same time, we consider that evidence does not negate a fault absolutely and let it down to a low level, here, 1% and according to formula (8), we can obtain its diagnostic entropy of each stage listed in Table 1, which have been normalized already.

The simple diagnostic process in Fig. 3 is coincident with a logarithmic model strictly. So here, we still use it to simulate inference process of engineering diagnosis.

$$H = \log (C_1j+C_2 ) \tag{9}$$

Then the decreasing ratio of diagnostic entropy is:

$$\frac{dH}{dj} \propto \exp (H_0 - H) \tag{10}$$

Another, sigmoid function is used in many fields, so we have:

$$H = \frac{1}{1 + e^{\alpha j + \beta}} \tag{11}$$

$$\frac{dH}{dj} \propto -H (1 - H) \tag{12}$$

Here,  $C_1$ (or  $\alpha$ ) and  $C_2$ (or  $\beta$ ) characterize the diagnostic process.  $C_1$  or  $\alpha$  represents efficiency of the inference

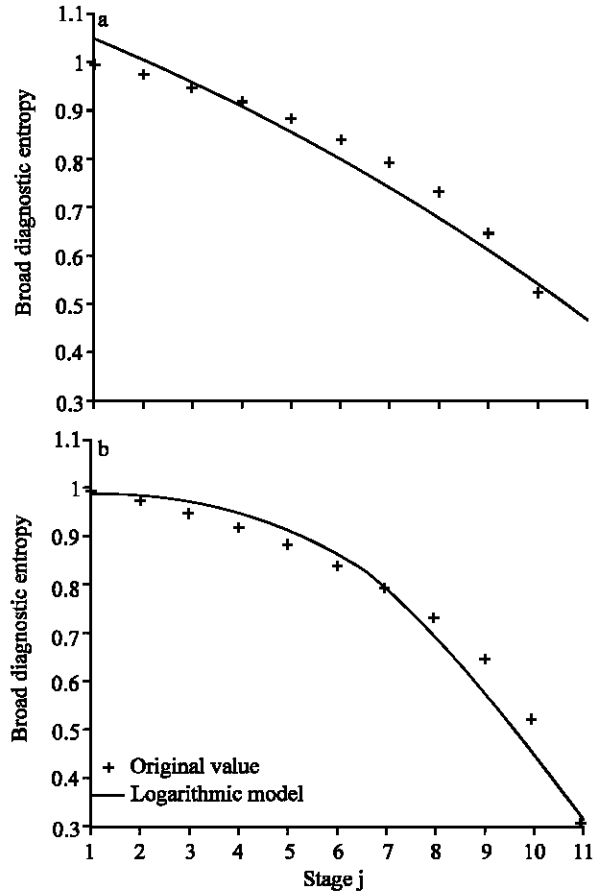


Fig. 4: A simple diagnostic process

process and  $C_2$  or  $\beta$  shows initial uncertainty of a diagnostic process. Figure 4 gives the results simulated with these two models.

From it, we can see that sigmoid function is better than logarithmic function. It means that, sigmoid model is better than logarithmic model to characterize the inference process of engineering diagnosis. We'll verify this by means of computer simulation afterwards.

Suppose there are N possible potential faults inside a running machine, we have to find the leading fault, which determines the machine operating condition. In engineering diagnosis, the obtained evidence negates one possible fault in each stage, which decreases its probability to a low level, such as 1%. Assume also that all the evidence obtained in each stage is with equal chance, i.e. all N-1 faults can be negated randomly in N stages and finally, there remains one leading fault to be

Table 1: Entropy in a simple diagnostic process

Stage	1	2	3	4	5	6	7	8	9	10	11
H	0.997	0.973	0.947	0.917	0.882	0.840	0.790	0.726	0.642	0.519	0.301

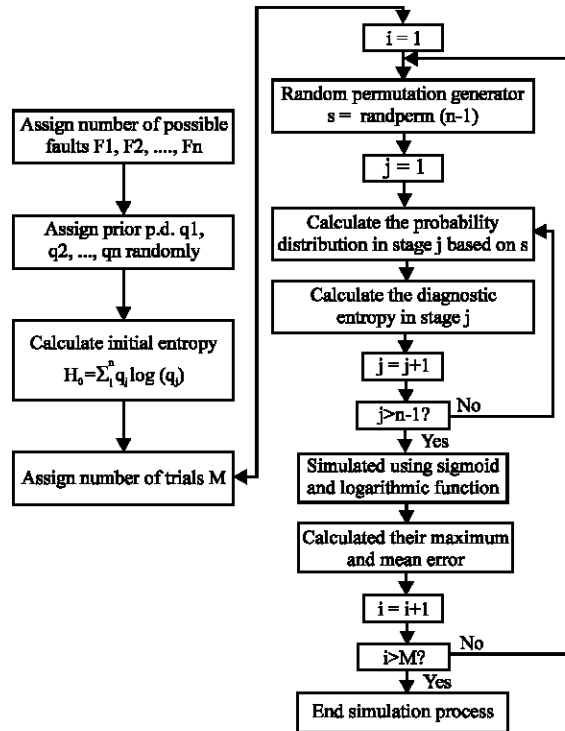


Fig. 5: The computer simulation algorithm

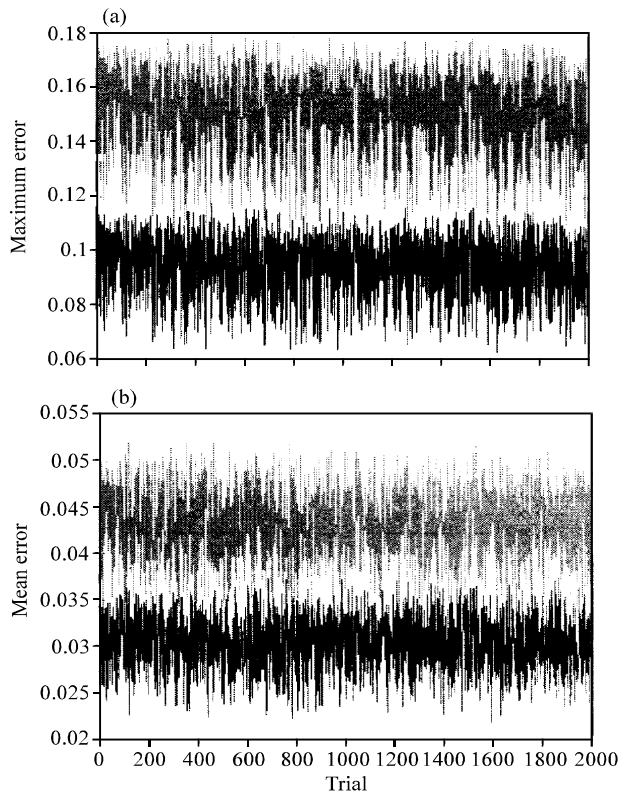


Fig. 6: The errors of computer simulation results

definitely diagnosed in the last stage. Then we can run computer to perform M trials (M is a large number, e.g. 5000). In each trial, a random number permutation generator  $S = \text{randperm}(N-1)$  is executed to generate a random order, e.g.  $S = [3, 1, 4, 6, 5, 7, 2]$  for  $N = 8$ . According to this random order, we can calculate the probability distribution in j-th stage as following:

$$p(j:N-1) = \frac{q[S(j:N-1)]}{1 - \sum q[S(1:j-1)]};$$

$$p(N) = \frac{q(N)}{1 - \sum q[S(1:j-1)]} \quad (13)$$

The diagnostic entropy in stage j can be calculated as following:

$$H_j = - \sum_j^N p(j) \cdot \log_{10} p(j) \quad (14)$$

for  $j = 2, \dots, N-1$ . For  $j = 1$ ,  $H = H_0$ , or the prior diagnostic entropy. Figure 5 shows the simulation algorithm.

Here, we perform  $M = 8000$  trials with the number of possible faults  $n = 12$ . The simulation results seem to be successful. Figure 6 shows the maximum and mean errors of sigmoid and logarithmic model. And errors of only 2000 trials are drawn. Among of them, the upper serials in two figures are errors of logarithmic model and the nether serials in two figures are errors of sigmoid model. So it shows that the latter model is better than the former one.

### A CASE STUDY

**Background knowledge:** An industrial gas turbine in an oil refinery of petroleum-chemical complex connects with a centrifugal compressor for reuse of exhaust gas. Its rigid rotor is supported in two elliptical bearings with an overhung disk. In June 2002, the rotor with its disk was sent to repair welding via laser beam due to the corrosion near the roots of blades. During the overhaul, a new rotor was assembled into the turbine instead of the original repaired rotor. After the overhaul, the overall vibration was too high and approached to  $70 \mu\text{m}$ . So in September, the factory once again shut down the turbine and the repaired rotor was once again put in use instead of the new rotor. But the p-p (peak to peak) values indicated by proxy meters still maintained on a high level ( $70 \sim 90 \mu\text{m}$ ).

Figure 7 is a 3-dimensional diagram of first harmonic component with four bearing sections. It simulates the vibration of rotor. Obviously, vibration level of the section 'A', which is the disk end of gas turbine, is the highest. We'd like to decide the root cause of this high vibration level.

Table 2: The probability distribution of possible faults in each stage

	Runout	Balancing status	Alignment	Bearing clearance	Permanent bending	Thermal bending	Rub	Other
Prior Prob.	0.050	0.300	0.200	0.130	0.100	0.100	0.100	0.02
E <sub>1</sub>	0.020	0.310	0.206	0.135	0.103	0.103	0.103	0.02
E <sub>2</sub>	0.020	0.020	0.298	0.195	0.149	0.149	0.149	0.02
E <sub>3</sub>	0.020	0.020	0.347	0.227	0.020	0.173	0.173	0.02
E <sub>4</sub>	0.020	0.020	0.418	0.273	0.020	0.020	0.209	0.02
E <sub>5</sub>	0.020	0.020	0.484	0.316	0.020	0.020	0.100	0.02
E <sub>6</sub>	0.020	0.020	0.650	0.150	0.020	0.020	0.100	0.02
E <sub>7</sub>	0.020	0.020	0.750	0.050	0.020	0.020	0.100	0.02

According to the past experience or the prior information, the possible malfunctions of industrial gas turbines chiefly are static mechanical and electrical runout, unbalance, misalignment, bearing problems, permanent bending, temporary thermal bending and rub between stationary and rotating part<sup>[5]</sup>. And there may have other unpredictable faults, for example, cotton yarn was left in the gas turbine once. Here, we suppose the probability of these unpredictable faults is 0.02 throughout the inference process. Their probabilities of occurring can be estimated as shown in the second row of Table 2. Among these faults, the rotor unbalance is the most probable fault. The diagnostic process regularly should give first priority to find its evidence.

From the conversation between the expert and the field practitioner (Appendix), we can withdraw the following evidences to identify the leading fault:

- E<sub>1</sub>: If the measured peak-to-peak values chiefly are due to pure mechanical and electrical runout, these values should not be so high. In general, they should be ranged below 0.02 mm.
- E<sub>2</sub>: The rotor together with half part of the coupling was balanced on balancing machine.
- E<sub>3</sub>: Vibration level was already high at low speeds and not influenced seriously by speed and load changes. That is to say, there was no possibility of existing temporal thermal bending of rotor.
- E<sub>4</sub>: The static bending of rotors was measured inside the tolerance limits before assembly.
- E<sub>5</sub>: Higher harmonics in the vibration spectra are small, which indicate that there exists light rub between rotor and stationary part of gas turbine.
- E<sub>6</sub>: The alignment between central lines of turbine and compressor does not obey the design requirements. Serious vibration caused by the rotor misalignment in operating condition and exhibited on FFT spectra high first harmonic component.
- E<sub>7</sub>: The bearing clearance approaches to the lower design limit.

The above listed evidences partly come from practice, or the object oriented knowledge, for example E<sub>2</sub>, E<sub>4</sub> and E<sub>7</sub>. The evidence E<sub>1</sub> simply comes from the expert's

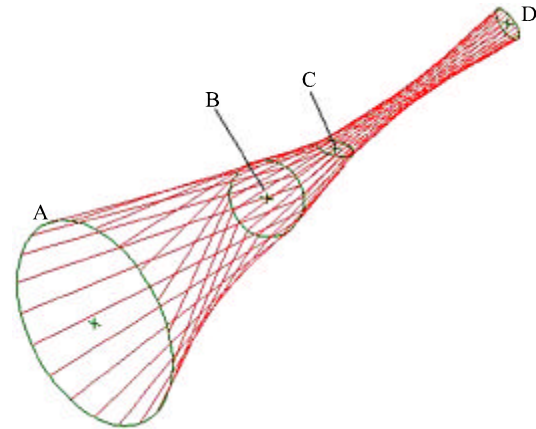


Fig. 7: The 3-dimensional diagram of first harmonic component with four bearing sections. Section A-the end of disk of the gas turbine rotor; Section B-the end of coupling of this rotor; Section C-the end of coupling of the compressor and the section D-another end of this rotor

quantitative conceptual knowledge. The evidence E<sub>3</sub> is also derived from the object oriented knowledge that the vibration amplitude did not obviously increase with the load applied. Regularly, the uneven rotor temperature is the root cause of its temporal thermal bending. The former is always directly proportional to the amount of inlet high temperature working medium and thus the applied load. These are the conceptual knowledge applied here by the expert. The evidence E<sub>5</sub> implies that the expert applies his methodological knowledge about FFT spectra and the symptom of rub failure on these spectra. Finally, the evidence E<sub>6</sub> is the result that the expert synthetically applied three kinds of diagnostic knowledge. First, in general, the case temperature of gas turbine in working condition is higher than that of centrifugal compressor it connected. Second, the thermal differential of two cases causes the misalignment between the two rotors and should be compensated in advance. That is to say, the central line of turbine rotor should be aligned a little lower than that of compressor rotor in non-operating condition. In practice, the former was higher than the latter. Third, the two rotors were connected by means of membrane coupling, not gearing coupling. So when the rotors are



Table 3: The kind of knowledge in each stage

Evidence	Kinds of knowledge
E <sub>1</sub>	Conceptual knowledge
E <sub>2</sub>	Object oriented knowledge
E <sub>3</sub>	Object oriented knowledge
E <sub>4</sub>	Object oriented knowledge
E <sub>5</sub>	Methodological knowledge
E <sub>6</sub>	OOK, CK, MK
E <sub>7</sub>	Object oriented knowledge

misaligned, the second harmonic component on the FFT spectrum is small, but first harmonic component high. Table 3 lists the kind of knowledge about the above evidence.

As the same with above, that certain fault is negated means its probability decrease to a low level, here, 2%. The change of probability values of each possible malfunction in Table 2 reflects the inference process in diagnosis. After conversation between field practitioner and expert, the expert received the information stream and finally, they found out the root cause of high vibration level.

In the case study listed above, there are eight possible root causes. The maximum information entropy is  $H_{max} = \log_2(8) = 3.0$  for full ignorance of any information. According to the past experience on the malfunction of industrial gas turbines, the expert gives an empirical distribution of these eight root causes. The entropy of this distribution is equal to 0.898 (Table 4), which is smaller than 3.0. In each stage of inference process the entropy gradually decreases until definite diagnosis. The probability distribution in each stage can be calculated using equations (8) and are listed in Table 2. These probability distributions are used to determine the information entropy  $H_0$ , as shown in the second row in Table 4. Figure 8a shows the entropy drop in the inference process using sigmoid and logarithmic model to simulate it.

In the above case study, the coefficients  $C_1$  and  $C_2$  equal to -0.12 and 2.64, respectively (equation 9) and  $\alpha=0.33, \beta=-2.58$  (equation 11). Their errors are 0.017 and 0.012 correspondingly.

Maybe if a wiser expert paid attention to the fact that neither the new rotor, nor the repaired rotor could improve the vibration condition, he could immediately comprehend that it would be due to other causes except unbalance or bending. Here, he could apply this fact as evidence and thus soon simplify the inference process. The new evidences are listed as following:

- E<sub>1</sub>: The p-p values due to pure mechanical and electrical run out should not be so high.
- E<sub>2</sub>: High vibration level both for new rotor and repaired rotor.

- E<sub>3</sub>: High vibration level at low speeds, which was not much influenced by applied load.
- E<sub>4</sub>: Small values of higher harmonics;
- E<sub>5</sub>: The alignment between the rotor central lines exceeds the design requirements.
- E<sub>6</sub>: The bearing clearance approaches to the lower design limit.

object-oriented knowledge, including the design features, the manufacturing and assembling process as well as the historical operation records, can be transformed to effective evidence. Indeed, the vibration behavior in run up stages and the type of rotor coupling played a decisive role in some situation. Therefore, the prerequisite of successful engineering diagnosis is to unite the equipment user with the manufacturer and the domain experts.

It is well known that the engineering diagnosis is a classification problem. One obvious difference between an expert and a novice in classification consideration

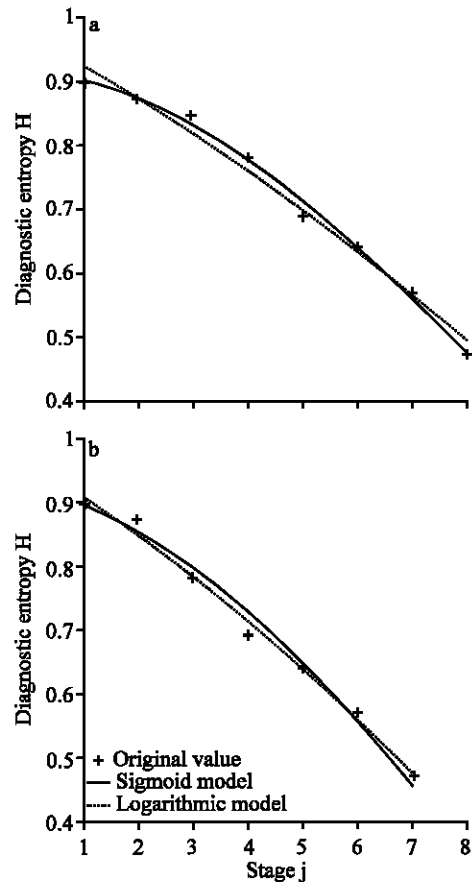


Fig. 8: the entropy drop of two case study: a) the first case study listed in the second; b) the another case which expert is more wiser

Table 4: The diagnostic entropy in two cases

j (Stage)	1	2	3	4	5	6	7	8
H <sub>0</sub>	0.898	0.874	0.849	0.781	0.691	0.643	0.570	0.475
H <sub>1</sub>	0.898	0.874	0.849	0.781	0.643	0.570	0.475	-

consists of feature extraction. As the experiments have revealed<sup>[6]</sup>, the domain expert is capable to catch the essence through the appearance and thus to support or negate some kinds of faults. In this case study, the fact that the vibration amplitude revealed no obvious increase in magnitude when the speed and load ran up is a surface phenomenon. It does not obey the fundamental principle of centrifugal force in physics, so the expert negated the rotor unbalance as the root cause. The temporal rotor bending is always accompanied by an uneven temperature rise and thus the load applied on the rotor, so the expert also negated it as the root cause. Above stated can be summarized as the typical ‘if-then not’ rules in constructing the expert system. In order to alleviate the dependence on experts and to estimate the prior probability distributions correctly, a sound knowledge base is absolutely necessary. It includes not only the historical records of individual machine set, typical case studies of similar equipment, but also the diagnostic expert system or artificial neural network. This can only be achieved under the cooperation of field practitioners and domain experts. The Bayesian diagnostic network is a powerful tool to represent the causal relationships among the symptoms and possible faults<sup>[7,8]</sup>. It can be applied in engineering diagnosis and consummated to improve the diagnostic process.

### CONCLUSION

Engineering diagnosis is essential to equipment safety, their continuous operation and predictive maintenance. As a key to successful engineering diagnosis, this study emphasized on the diagnostic knowledge, which are the conceptual knowledge, the methodological knowledge and the object oriented knowledge and examined the knowledge structure according to its role in diagnosis. Some possible measures for improving the diagnostic knowledge in engineering diagnosis were also mentioned.

Information entropy is a system uncertainty measure, which can be used in engineering diagnosis, named diagnostic entropy. Based on it, the study introduced sigmoid model and logarithmic model to simulate diagnostic process. Several examples, computer simulation and the diagnostic practice in an oil refinery revealed that sigmoid model is more efficient than logarithmic model to simulate the inference process of engineering diagnosis.

### Appendix: Conversation between expert and practitioner:

- P. We suspect the high vibration level is due to the initial mechanical and electrical runout of rotor measuring surface.
- E. Were these two rotors balanced in advance?
- P. Yes, these two rotors were balanced on a balancing machine to a high accuracy before mounting. We hope to measure the runout of the rotor.
- E. In general, the rotor runout can't approach to such a high peak-to-peak value, at most in the range of 10~20 μm (E<sub>1</sub>). When you balance a rotor on a balancing machine, you should put the rotor together with the coupling into consideration.
- P. Yes, the rotor together with half of the coupling was put on the balancing machine (E<sub>2</sub>). If the runout isn't playing a good role, why the peak-to-peak value has already risen to more than 60 μm in the run-up stages even when the speed was as low as 1000 rpm. At rotating speed under full load, the peak-to-peak value has raised to more than 70 μm (E<sub>3</sub>).
- E. This phenomenon tells us that the root cause is neither unbalance, nor temporary thermal bending (The difference in vibration amplitudes are relatively small when the rotor operated in different speeds and under different loads).
- P. In the run-up stages, the p-p values were recorded as shown in Table 5.
- E. There may be two possible causes: The rotor is with excessive permanent bending, otherwise, there may exist bearing faults, for example, excessive bearing clearance or improper temperature of the lubricating oil, etc.
- P. The static bending of the rotor was measured after repair. Its value is located in the tolerance limits. So it should not be a rotor with permanent bending (E<sub>4</sub>).
- E. There are small higher harmonics on the spectra as shown in Fig. 9. So there may be a light rub between rotor and stationary part (E<sub>5</sub>). What is the alignment condition between turbine and compressor?
- P. The coupling is of a flexible membrane type. According to design requirement, in static condition the central line of gas turbine should be lower than that of compressor by 0.1~0.2 mm, in order to compensate the thermal expansion differential between the two cases. In this machine set the central line of turbine is even higher than compressor in cold condition (E<sub>6</sub>).

Table 5: The vibration p-p values in run-up stages of the industrial gas turbine

Rotor speed (rpm)	Vibration p-p value near bearing sections ( $\mu\text{m}$ )	
	Disk end	Coupling end
1000	70~73	15~20
2800	70~73	15~20
4475	70~76	12~20
5254	74~78	10~20
5887 (no load)	75~78	10~20
5887 (full load)	77~90	10~20
5887 (stable operation)	73~84	10~20

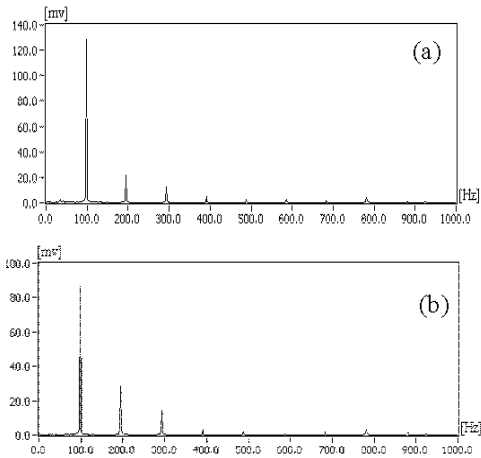


Fig. 9: The FFT vibration spectra near the bearing section at disk end

E. When a coupling of membrane type connects two rotors, the misalignment exhibits a high 1x harmonic vibration with very small 2x harmonic. This symptom is different from the misalignment between shafts connected by means of gear coupling. That is why the two rotors all exhibited high peak-to-peak values even at low speeds and with only a little increase at higher speeds and finally it can still maintain in stable operation up to now. How did you set the bearing clearance?

P. The bearing clearance is 0.18 mm. This value approaches to the lower design limit. ( $E_7$ ) We suppose to further shrink this clearance.

**REFERENCES**

1. Robert, J.D., J.L. William and J.G. William, 2002. Model of knowledge, cognition and learning. Umass Physics Education Research Group. <http://umperg.physics.umass.edu/perspective/CognitiveModelPaper>, 2002.
2. Kullback, S., 1969. Information Theory and Statistics. New York: Dover.
3. Shore, J.E. and R.W. Johnson, 1981. Properties of cross entropy minimization. IEEE Trans., IT-27: 472-482
4. Heylighen, F. and C.I. Joslyn, 2001. Entropy and information. Principia Cybernetica Web, <http://pcp.lanl.gov/ENTROINFO.html>. 2001
5. Qu, L.S. and G.H. Xu, 1999. One decade of holospectral technique: Review and prospect. DETC99/RSAFP-8850, Proceedings of the 1999 ASME Design Engineering Technical Conference, Las Vegas, Nevada, September, pp: 12-15.
6. Michelene, T.H. Chi, P.J. Feltovich and R. Glaser, 1980. Categorization and representation of physics problems by experts and novices. Cognitive Sci., 5: 121-132.
7. Langley, P., W. Iba and K. Thompson, 1992. An analysis of Bayesian classifiers. Proceedings of the 10th Natl. Conf. Artificial Intelligence, AAAI Press, pp: 223-228.
8. Nir, F., 2004. Learning bayesian networks from data. U.C. Berkeley. <http://www.cs.berkeley.edu/~nir/tutorial>.