An Intelligent Integrated System Scheme for Machine Tool Diagnostics

Wenbin Hu*, Andrew Starr*†, Zude Zhou** Andrew Leung*

*Manchester School of Engineering, University of Manchester, Manchester, M13 9PL, UK

**School of Mechanical Science & Engineering, Huazhong University of Science & Technology, Wuhan, 430074, P.R. China

† To whom correspondence should be addressed

Abstract: The technology of neural networks and expert systems are finding increasing applications in the field of machine tool diagnostics. In this paper, the advantages and disadvantages of these methods are analysed and compared. An intelligent integrated diagnosis system based on a combination of the two methods is presented. This scheme aims at exploiting the advantages and avoiding the disadvantages of neural networks and expert systems. The implementation of the intelligent integrated diagnosis system scheme is also presented. A diagnosis system based on the scheme is introduced, which was applied to the process diagnosis of an existing machining centre. The experimental results show that the integrated system scheme is feasible and effective for machine tool diagnosis tasks.

Keywords: machine tool, fault diagnosis, neural networks, expert systems, integration.

1. Introduction

In recent years, the technology of machine tool diagnostics has made great strides forward. In particular, with the rapid development of artificial intelligence, many powerful intelligent diagnostic techniques have been made available to meet the needs of diagnosing modern complex machine tools. Neural networks and expert systems are two popular methods and have been used independently in practice with varying degrees of success.

However, neural networks and expert systems are two quite different approaches to diagnostics. They have different properties as well as advantages and disadvantages with regard to the diagnosis of faults [1]. Neural networks are based on numeric computations and algorithms, while expert systems are based on symbolic and heuristic reasoning. Neural networks have capabilities of association, memorisation, error-tolerance, self-adaptation and multiple complex pattern processing. On the other hand, they cannot explain their own reasoning behaviour and cannot diagnose new faults (those not already made available previously in training the networks). Expert systems have obvious knowledge representation forms that make knowledge easy to manage. Compared with neural networks, expert systems have the ability to explain their reasoning behaviour and can diagnose new faults using knowledge bases. However, self-learning is still a problem and computation time can be lengthy for difficult diagnostics tasks. It is therefore sensible to combine neural networks with expert systems for machine tool diagnostics as the advantages of one approach can outweigh the disadvantages of the other. This scheme can be effective in modern machine tool fault diagnosis.

Similar ideas were reported about combining neural networks and expert systems in the application to areas such as computer security [2] and medical diagnosis [3]. In the majority of cases, neural networks and expert systems were developed separately with one way data passages from neural networks to expert systems, and with interaction between the user and the two models. Some of them permit both components to receive data concurrently. In this paper, a new scheme for integrating neural networks and expert systems is introduced for machine tool fault diagnosis. In this scheme the two models are completely integrated as a single unit. The integrated system has been implemented and simulated on an existing machining centre.

2. Machine Tool Diagnosis Based on Expert Systems

Expert systems have been adopted extensively for machine tool diagnosis. They employ expert reasoning methods and computer models to solve problems, and draw similar conclusions to those derived by experts. Using expert systems, diagnostic goals can usually be achieved by specific knowledge representation forms and problem solving techniques, and machine fault sources can be deduced in conjunction with machine conditions. An expert system is composed of a database, a knowledge base and a reasoning engine. When diagnosing a machine tool fault, the reasoning engine carries out reasoning under the guidance of reasoning mechanisms, according to the fault data in the database, and diagnostic knowledge in the knowledge base. A typical machine tool diagnosis expert system is shown in Figure 1.

Although expert systems have been applied to machine tool diagnosis with some success, problems have been experienced, predominantly due to the following limitations:

- knowledge acquisition is difficult; this acquisition is one of the major bottlenecks in developing expert systems, and requires long periods of time and is costly [4];
- it is difficult to completely mimic the human thought process [5];
- the knowledge base is never complete; when a new fault is encountered, to which no formal knowledge is available, expert systems become ineffective [6].

One way of overcoming these limitations is to improve the system’s ability for automatic knowledge acquisition, self-learning, association and memorisation. This is where the strengths of neural networks can be incorporated.

3. Machine Tool Diagnosis Based on Neural Networks

Traditional neural network techniques may be used to diagnose faults in relatively straightforward simple machines with single processes, single faults or gradually occurring faults. However, modern machine tools usually operate multiple processes, and manifest multiple faults, some of which can be catastrophic. Moreover, these machines are complex and highly automated.

The multi-layer perceptron (MLP) neural network, trained with the back propagation (BP) algorithm, is the most commonly used for machine fault diagnosis [7]. MLP network based machine diagnosis is usually performed as follows.

- A feature space (input space) $X=[x_1, x_2, \ldots, x_n]^T$ is constructed, where x_i is the observed value of the i-th feature. These feature values are obtained through signal analysis and processing during the operation of the machine. The detailed procedure includes signal measurement, signal processing and feature extraction.
- A class set $C=\{c_1, c_2, \ldots, c_k\}$ is defined for each given set of machine conditions, and an output space $Y=[y_1, y_2, \ldots, y_m]^T$, where c_i is the i-th class and y_j is the j-th output respectively. A specific class is represented through the output space Y.
- The neural network is constructed as indicated in Figure 2 according to the input and output spaces defined above, and the hidden layers and their neurons are selected. The number of input neurons is equivalent to the number of the features selected. The number of output neurons is equal to the dimension of the output space.
- The parameters in the networks such as weight ω^0_{11} are estimated by training, after a set of samples are obtained. In other words, if $S=\{s_1, s_2, \ldots, s_k\}$ is a sample set of a specific class c_i, the neural networks are trained through every pair of input-output (s_i,

c_i), where i=1, 2, …, m;

- The decision constraint is established automatically after the training process has converged. The neural networks are then able to classify machine conditions for any input $x_n \in X$.

The selection of the activation function is also important. The most frequently used function is the sigmoid function, and is given by:

$$f(x) = \frac{1}{1+e^{-x}} \tag{1}$$

where x is the sum of inputs of a certain neuron and $f(x)$ is the output of the neuron.

To use neural networks to diagnose machine tool faults, there must be a corresponding output neuron for each fault class. Therefore, when a new fault not available previously during the training process is presented, the diagnosis system is incapable of diagnosing the fault. In addition, neural networks cannot explain their own reasoning behaviour, but this can be supplemented using expert system.

4. The Integration of Neural Networks and Expert Systems

In developing expert systems for machine fault diagnosis, the acquisition of diagnostic knowledge usually becomes a major bottleneck due to:

- the indirect method of interviewing domain experts who would use “common sense” or anecdotal evidence;
- insufficient training, making it difficult for engineers to interview domain experts efficiently and effectively, and for programmers to codify their knowledge;
- little or no expert knowledge available in many scenarios.

Many algorithms or models have been employed to automatically extract diagnostic knowledge from training samples. The automatic generation of knowledge bases saves time, expense and human resources and can be applied to situations with only data but no human expert being available. To improve the accuracy of diagnosis, an approach has been created which automatically generates the knowledge base using neural networks, and integrates neural networks with expert systems [8]. The design of such an integrated system is presented in Figure 3.

In Figure 3, the meta-system is designed to carry out overall control of the integrated system. It acts as a man-machine interface as well as a data collector. It can also call neural networks to make a diagnosis, and explain and evaluate the results of the diagnosis. When the neural networks are unable to make a diagnostic decision on a new fault because of a lack of data, the meta-system will activate the expert system to diagnose the fault by using deep knowledge. Deep knowledge is defined as knowledge about the machine structure, behaviour and function, the acquisition and representation of which will be introduced in section 5.3. In this case, the meta-system acts as a data classifier. At run time, it classifies incoming data or symptoms and sends the results to the neural networks or expert system for further diagnosis. The final diagnostic result is taken as a fault sample which is used to train the neural network. The learning mechanism is then initiated by the meta-system to extract new diagnostic knowledge from the training sample.

Based upon the hierarchical composition of the machine, neural networks diagnose machine faults hierarchically to guarantee prompt diagnosis. The hierarchical structure of neural networks is arranged as shown in Figure 4. At run time, the fault locating network classifies incoming data or events, then sends the results to the networks in the second layer and calls corresponding sub-networks to perform further diagnosis. The networks in the second layer include all the sub-networks associated with the individual components of the machine. These sub-networks are responsible for fault diagnosis of the related component. They are all two-layer (input, hidden neurons and output neurons) MLP networks.

The training of the neural networks can be carried out in two ways, namely:

- Manual training for common faults - the training can be carried out by users through the man-machine interface.
- Automatic training for new faults (those not included in previous training samples) - the training is conducted automatically by the system itself.

In the latter case, the meta-system summons the expert system to initiate deep diagnosis and start the learning mechanism to train samples based on the diagnostic results. This process can enrich the diagnostic knowledge and enhance the diagnostic ability of the system.

Such an integrated system exploits the advantages and avoids the disadvantages of neural networks and expert systems, whose mutually exclusive properties are summarised below:

Neural networks:

- Association
- Memory
- Error-tolerance
- Self-learning
- Self-adaptation
- Processing multiple complex patterns

Expert systems:

- Able to explain reasoning behaviour
- Diagnose new faults using diagnostic knowledge

5. Implementation of the integrated system

5.1 Fault data collection and training samples preparation

To diagnose fault characteristics accurately in complex modern machines, multiple parameters or signals must be collected and analysed. Generally these parameters or signals can be classified into three categories:

- machine and process status, which is collected by a multi-sensor and multi-parameter based condition monitoring system;
- signals in machine controllers, which can be acquired through an I/O interface between machine controllers and the diagnosis system computer;

- symptoms observed by machine operators, which are entered to the diagnosis system database through the users' interface.

Training samples are obtained by representing experiential knowledge in a numerical format. The experiential knowledge is previously acquired from experiences, fault history and theoretical fault analysis provided by maintenance personnel for each machine. When preparing samples those typical fault cases and valid fault signs, which are critical to the success of the diagnostic networks, must be selected.

5.2 Learning algorithm

Error back propagation (BP) is the most widely used learning algorithm since it is relatively straightforward to implement and, most importantly, often outperforms other methods. The BP algorithm is routinely used to train a MLP network. Its learning procedure consists of both forward-propagation and back-propagation. During forward-propagation, all information is entered at the input layer and processed at the hidden layers, and finally transferred to the output layer. The status of neurons at each layer only affects the status of those at the next layer. If the expected output cannot be obtained at the output layer, it will conduct back-propagation. During back-propagation, the error from the output layer is transmitted back, through the hidden layers, to the input layer. In this procedure, the weights at every layer are updated to minimise the error (a previously defined value) by adapting the weights in such a way that the energy function is minimised using gradient descent.

Suppose there are N training samples, i.e. N input-output pairs (X_k, Y_k), k=1, 2, …, N, where $X_k=\{x_{k1}, x_{k2}, \ldots, x_{kn}\}$ is the input vector of the k-th sample, and $Y_k=\{y_{k1}, y_{k2}, \ldots, y_{km}\}$ is the output vector of the k-th sample. Here $\underline{Y}_k=\{\underline{y}_{k1}, \underline{y}_{k2}, \ldots, \underline{y}_{km}\}$ is defined as the real output of the output layer. $O^p{}_k=\{o^p{}_{k1}, o^p{}_{k2}, \ldots, o^p{}_{kl}\}$ the output of p-th hidden layer, and ω_{ij} the connection weight from the i-th input to the j-th output. When training the network using the k-th sample, the error is defined as

$$E_k = \frac{1}{2}\sum_{i=1}^{m}(\underline{y}_{ki} - y_{ki})^2 \tag{2}$$

and the total error $E = \sum_{k=1}^{N} E_k$. The weight ω and threshold θ (of each node) are updated according to

$$\begin{aligned} \omega_{ij}(k+1) &= \omega_{ij}(k) + \Delta_k \omega_{ij} \\ \theta_j(k+1) &= \theta_j(k) + \Delta_k \theta_j \end{aligned} \tag{3}$$

where $\Delta_k\omega_{ij}$ and $\Delta_k\theta_j$ are the correction values of weight and threshold value respectively. They are expressed by

$$\begin{aligned} \Delta_k \omega_{ij} &= \mu\delta_{ij} O_{ki} \\ \Delta_k \theta_j &= \mu\delta_{kj} \end{aligned} \tag{4}$$

where μ (>0) is the learning velocity (i.e. the step length for gradient searching), and δ_{kj} the training error of the k-th sample at the j-th layer.

The detailed training steps are as follows.

(1) Initialise ω_{ij} and θ_j, i.e. set the values of $\omega_{ij}(0)$ and $\theta_j(0)$.

(2) Provide training samples, i.e. input vector X_k (k=1, 2, …, N), and objective output vector Y_k (k=1, 2, …, N), and to carry out the following steps (3)~(5) for each sample.

(3) Calculate the real output and status of the hidden layers (forward-propagation). For the k-th sample, the output of the p-th layer of the network is

$$O_{kj}^p = f(net_{kj}) = f(\sum_i \omega_{ij}^p O_i^p + \theta_j^p) \tag{5}$$

(4) Calculate the training error. For the BP networks with a sigmoid activation function, the training error at the output layer is

$$\delta_{kj} = O_{kj}(1 - O_{kj})(\underline{y}_{kj} - y_{kj}) \tag{6}$$

and at the hidden layers,

$$\delta_{kj} = O_{kj}(1 - O_{kj})\sum_{l}\delta_{kl}\omega_{lj} \tag{7}$$

(5) Update weight ω_{kj} and threshold θ_j by

$$\begin{aligned} \omega_{ij}(k+1) &= \omega_{ij}(k) + \mu\delta_{kj}O_{ki} \\ \theta_j(k+1) &= \theta_j(k) + \mu\delta_{kj} \end{aligned} \tag{8}$$

(6) After training N times, check whether the error has satisfied the precision requirement, i.e., $E<\varepsilon$ (ε is a small positive number), and if yes, to stop training, or to return to step (3) till $E<\varepsilon$.

5.3 Knowledge acquisition and diagnostic reasoning

Both deep knowledge and shallow knowledge (experiential knowledge) are needed in the diagnosis of machinery faults. The former consists of knowledge about the machine structure, behaviour, function, fault tree analysis and parameter/state estimation. The latter is predominantly obtained from engineers' and maintenance experts' experience, fault statistics, process history or extracted from the fault samples by neural networks. Knowledge acquisition may be described as in Figure 5. This knowledge is coded and saved in a comprehensive knowledge base, which contains deep knowledge as well as shallow knowledge. Deep knowledge is represented as frames or rules, while shallow knowledge is represented as rules or independent facts. The modelling is further described in [9].

A high-quality diagnostic strategy must guarantee the efficiency of diagnosis, and in the meantime achieve good results. From this point of view, the integrated diagnostic scheme combines neural networks and expert systems into one unit, with the former guaranteeing diagnosis efficiency and the latter guaranteeing good diagnosis results. When there is a fault, the integrated system first calls the appropriate neural network for diagnosis. If it produces no result because the fault is not yet trained into the network, the expert system will be activated to complete the diagnosis using the knowledge base. The neural networks and the expert system cooperate with each other in order to locate a fault. Once the neural networks are trained, the diagnosis using the integrated system can be performed in the procedures as shown in Figure 6.

6. Case Study

The integrated diagnosis scheme has been implemented in a fault diagnosis system for an existing PFZ1500 machining centre. The machining centre is part of a flexible manufacturing cell, which is used at the Zhengzhou Textile Machinery Plant in China, for making machine parts. In the diagnosis system, the machine condition is monitored through measuring multiple parameters such as vibration, temperature, power and pressure. The monitored components include spindle, spindle motor, tool/work-piece, X axis, Y axis, Z axis, the drive motors in the X, Y and Z axes, mains power, and pneumatic and hydraulic sources. According the monitoring results, faults can be located in these areas.

For the PFZ1500 machining centre, 45 samples were collected in total. The samples and output modes for each monitoring component are listed in Table 1.

The diagnosis program was developed using Visual C++. Various parameters about network structure, initial weights and training samples were set before training. The learning velocity of the networks was set at 0.8, i.e., μ = 0.8 (refer to Equations (4) and (8)). The criterion to end training was that the error between every real output and its ideal output did not exceed 0.02, i.e., $E \leq 0.02$ (refer to Equation (2)). Comparing several learning results, the number of neurons in the hidden layer was set to 10. Meanwhile, the numbers of input and output neurons were set to be 45 and 12 respectively according to the total number of fault samples and output modes shown in Table 1. In addition, the initial weights ($\omega_{ij}(0)$) could be produced randomly between −0.4 and +0.4. The test results have shown that the threshold θ of an output can be set to 0.05. This means that when an output value Y_k lies in the range of [0.05, 1], it indicates that a fault related to the output neuron has occurred. On the other hand, when an output value lies in the range of [0, 0.05], it indicates that there is no fault of this nature.

The deep knowledge base contained the knowledge about the structure and behaviour of the machining centre, which was used to diagnose faults not previously available in the network training samples. The deep knowledge was represented in object-oriented classes. A total of 72 classes were used in the deep knowledge base. The meta-system contained knowledge about control and evaluation, which are represented as rules and facts. In total 48 rules and 30 facts were used in the meta-system.

The integrated diagnosis system was simulated separately for most kinds of faults in the 12 possible fault areas. If we define P as the rate of correct diagnosis, N_1 the number of correct diagnosis and N_2 the number of false diagnosis, then we have

$$P = \frac{N_1}{N_1 + N_2} \tag{9}$$

In our study, the results of experiments produced an average value for P of greater than 91.5%. Further detail of the application can be found in [10].

7. Conclusion

The integration of neural networks and expert systems into machine fault diagnosis takes full advantage of the low-level processing capability of neural networks and the high-level reasoning capability of expert systems. Because neural networks can ensure reliability and expert systems are able to guarantee completeness, the integration of the two techniques guarantees highly reliable and complete diagnosis. The combined technique is particularly suitable for real-time machine fault diagnosis. A diagnosis system based on the technique have been developed, and the system has been on trial run since 1997 on a PFZ1500 machining centre and good results have been achieved so far.

References

[1] A. H. Zhang, Monitoring and diagnostic techniques for mechatronic equipment, Press of Northwest University of Technology, Xi'an, 1995.

[2] S. Limbach, K. Miller, Neural nets, expert systems and computer security, Proc 11th Annual Midwest Computer Conference, 1997.

[3] E. Prem, Integrating knowledge-based components into neural network software, Österreichisches Forschungsinstitut für Artificial Intelligence, Wien, TR-95-18, 1995.

[4] A. K. Ray, Equipment fault diagnosis - a neural network approach, Computers in Industry 16, 1991, pp. 169-177

[5] W. Hu, Research on Integrated Fault Detection and Diagnosis Strategies as well as System Approaches for Modern Manufacturing Equipment, Post-doctoral Research Report, Huazhong University of Science & Technology (Wuhan), 1995.

[6] R. Davis, Expert systems: where are we? and where do we go from here?. AI Magazine 1, 1982.

[7] Y. B. Chen, X. Li, E. Orady, Integrated diagnosis using information-gain-weighted radial basis function neural networks, Computers in Industrial Engineering 2, 1996

[8] Z. Wang, Y. Liu, P. J. Griffin, Combined ANN and expert system tool for transformer fault diagnosis, IEEE Engineering Society, Winter Meeting, Vol.1, 1999, pp. 339-347

[9] Hu W, Starr A G, Leung A Y T, 1999, Two diagnostic models for PLC controlled flexible manufacturing systems, International Journal of Machine Tools & Manufacture, Vol 39 Iss 12 pp1979-1991, ISSN 0890-6955

[10] Hu W, Starr A G, Leung A Y T, 2000, A systematic approach to integrated fault diagnosis of flexible manufacturing systems, International Journal of Machine Tools & Manufacture, ISSN 0890-6955 v40 pp1587-1602

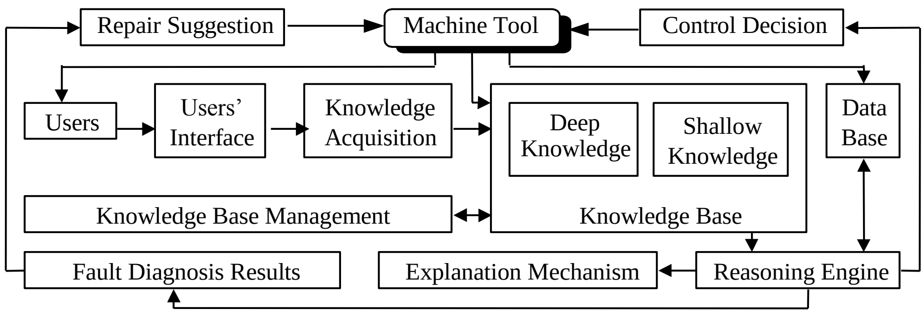

Figure 1. Machine tool diagnosis expert system

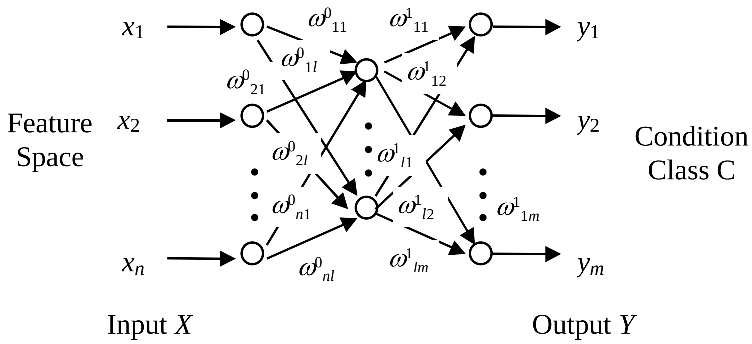

Figure 2. BP neural network based diagnosis model

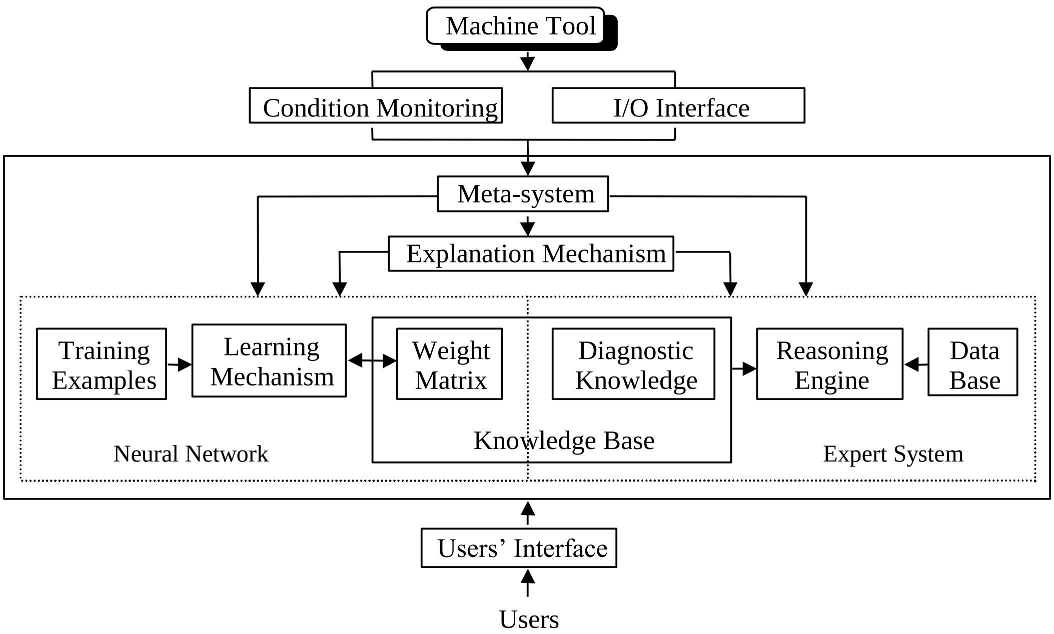

Figure 3. An integrated machine tool fault diagnosis system

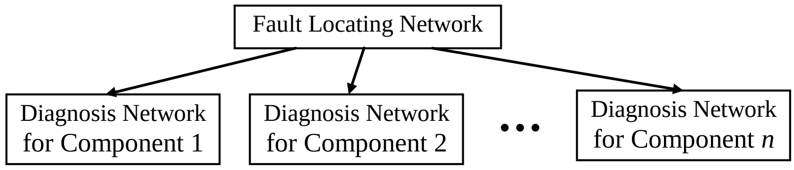

Figure 4. The hierarchical structure of neural networks

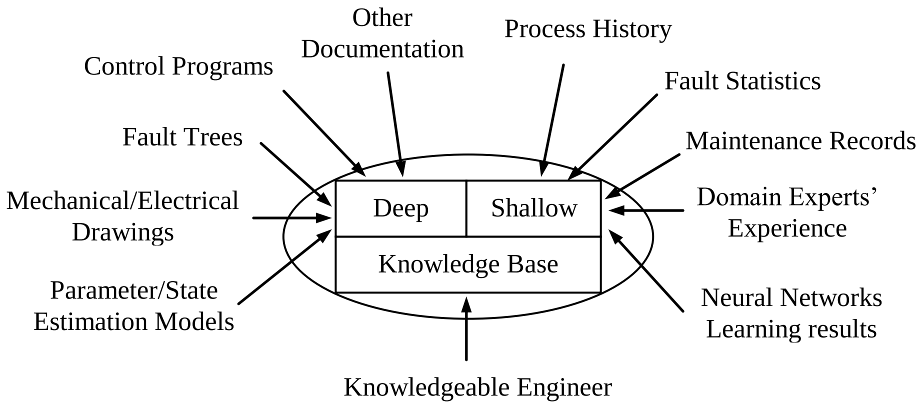

Figure 5. Diagnostic knowledge acquisition

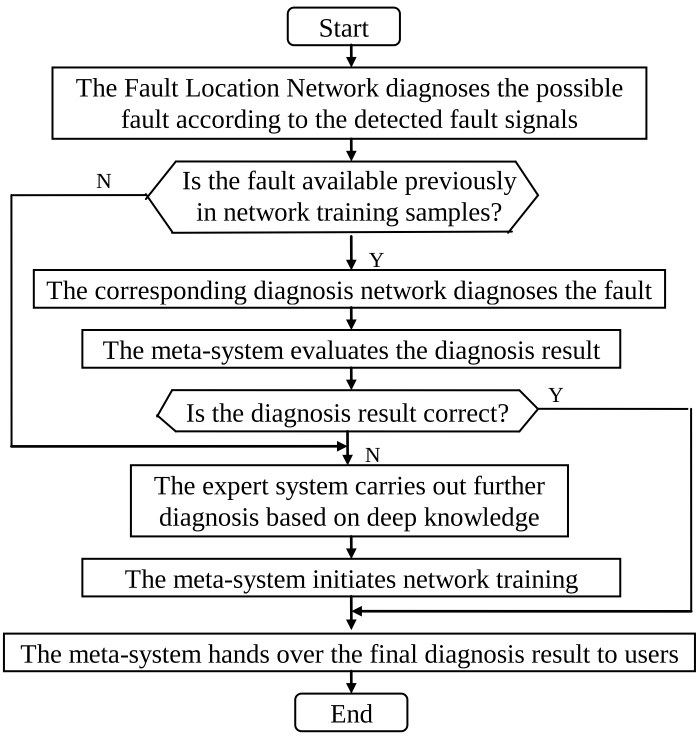

Figure 6. The diagnosis procedure of the integrated system

Table 1. Samples of process faults for the PFZ1500 machining centre

Fault Areas	Sample Number	Output Modes											
		y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}
spindle	6	1	0	0	0	0	0	0	0	0	0	0	0
spindle motor	3	0	1	0	0	0	0	0	0	0	0	0	0
tool/work-piece	7	0	0	1	0	0	0	0	0	0	0	0	0
X axis	4	0	0	0	1	0	0	0	0	0	0	0	0
Y axis	4	0	0	0	0	1	0	0	0	0	0	0	0
Z axis	4	0	0	0	0	0	1	0	0	0	0	0	0
X axis drive motor	3	0	0	0	0	0	0	1	0	0	0	0	0
Y axis drive motor	3	0	0	0	0	0	0	0	1	0	0	0	0
Z axis drive motor	3	0	0	0	0	0	0	0	0	1	0	0	0
mains power	2	0	0	0	0	0	0	0	0	0	1	0	0
pneumatic pressure	3	0	0	0	0	0	0	0	0	0	0	1	0
hydraulic pressure	3	0	0	0	0	0	0	0	0	0	0	0	1