An Approach to Fault Diagnosis of Chemical Processes via Neural Networks

J. Y. Fan, M. Nikolaou, and R. E. White
Dept. of Chemical Engineering, Texas A&M University, College Station, TX 77843

This article presents an approach to fault diagnosis of chemical processes at steady-state operation by using artificial neural networks. The conventional back-propagation network is enhanced by adding a number of functional units to the input layer. This technique considerably extends a network's capability for representing complex nonlinear relations and makes it possible to simultaneously diagnose multiple faults and their corresponding levels in a chemical process. A simulation study of a heptane-to-toluene process at steady-state operation shows successful results for the proposed approach.

Introduction

Application of artificial neural networks to fault diagnosis relies on their ability to "learn" and perform nonlinear mappings and, thus, to recognize patterns. Performance patterns for chemical processes can be represented by their measured state variables. Any fault or malfunction in a process may result in an abnormal performance pattern. The objective of fault detection, thus, is to recognize such abnormal patterns and to find the corresponding causes of faults or malfunctions.

Fault diagnosis for chemical processes via neural networks has elicited notable research interest recently (Hoskins et al., 1988). Venkatasubramanian and Chan (1989) reported a study in which a neural-network-based methodology was compared to a rule-based detection technique. A fault-tree-based knowledge representation was utilized which was then encoded into binary patterns presented at the neural network. The results of the study indicated that the neural-network-based version was better than the conventional rule-based one in several aspects. Watanabe et al. (1989) presented a study for detecting single faults in a heptane-to-toluene process. A two-stage structure of a neural network was proposed, with the first stage used to detect the fault type and the second one to detect the corresponding severity level. The network topology used in each stage was the same, with the input patterns being encoded in continuous values and the outputs in binary-like values. This configuration, however, has an inherent inability to tackle multiple faults and is inconvenient for real-time implementation, since it requires many subnetworks.

Correspondence concerning this article should be addressed to R. E. White.

Motivation and Significance

In chemical processes, the relationships between performance patterns and fault causes are generally nonlinear. In multiple fault cases, these relationships can be highly nonlinear and complex and, thus, may significantly exceed the inherent representation capability of a conventional back-propagation network. This may cause unacceptable training errors and wrong detections to occur. Similar symptoms may also appear even in single fault cases, if simultaneous identification of fault types and their corresponding levels via a single network is attempted. In this study, we present an approach that attempts to resolve the above issues based on modified back-propagation neural networks. More specifically, this approach addresses the following two problems:

- Effective detection of multiple faults occurring simultaneously.
- Simultaneous identification of fault types and their severity levels.

The modification we propose alters the configuration of a conventional back-propagation network by adding to the input layer a number of functional units that perform orthonormal transforms on individual inputs. This modification does not require new information from process measurements, but successfully expands the input pattern into a large space and, therefore, enhances the representation capability of the network.

The effectiveness of our proposed approach is examined through simulation of fault detection for a heptane-to-toluene aromatization process in steady-state operation (Watanabe et al., 1989). Successful detections for various fault cases, in-

cluding single and double faults at different levels, are obtained. The fault types and their corresponding levels are simultaneously identified. Simulations also demonstrate that the modified network can learn from typical representative fault patterns to perform general detections for possible faults under consideration.

In the rest of the article we first present an augmented neural network, and in the subsequent section we present extensive simulations. We draw conclusions in the final section.

A Modified Neural Network

Investigations have shown that back-propagation networks possess the ability to learn and perform nonlinear mappings, but they can only do that well to a certain extent. For the case of fault detection of chemical processes, nonlinearities can be very complex, especially in the case of multiple faults with different severity levels. Representation of such nonlinearities can exceed the capability of a conventional back-propagation network. This can be conceptually shown as follows:

Consider a nonlinear function

$$z=f(x_1, x_2, \ldots, x_n) \tag{1}$$

which represents an n dimensional hypersurface. Assume that only those values of the function at m points in the surface are of interest. If $m \leq n$, the nonlinear relationships between z and the vector $x=[x_1, x_2, \ldots, x_n]$ at the m points can be equivalently represented by a two-layered network and can be mathematically expressed by (Pao, 1989):

$$y^{(i)}=\frac{1}{1+\exp[-\beta z^{(i)}]} \tag{2}$$

$$z^{(i)}=\text{net}^{(i)}-\theta \tag{3}$$

$$\text{net}^{(i)}=\sum_{j=1}^{n} w_j x_j^{(i)}, \quad i=1, 2, \ldots, m \tag{4}$$

and

$$\begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \cdots & x_n^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(m)} & x_2^{(m)} & \cdots & x_n^{(m)} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} z^{(1)}+\theta \\ z^{(2)}+\theta \\ \vdots \\ z^{(m)}+\theta \end{bmatrix} \tag{5}$$

or, concisely,

$$XW=Z$$

Equation 5 is a set of linear equations that may have at least one solution for the weight vector $W=[w_1, w_2, \ldots, w_n]$. However, when $m>n$, no vector W, in general, will exist to satisfy Eq. 5. A solution

$$W=(X^TX)^{-1}X^TZ \tag{6}$$

from pseudo-inversion may give a best fit to which the back-propagation learning may eventually converge, but high error values may appear at the end of learning.

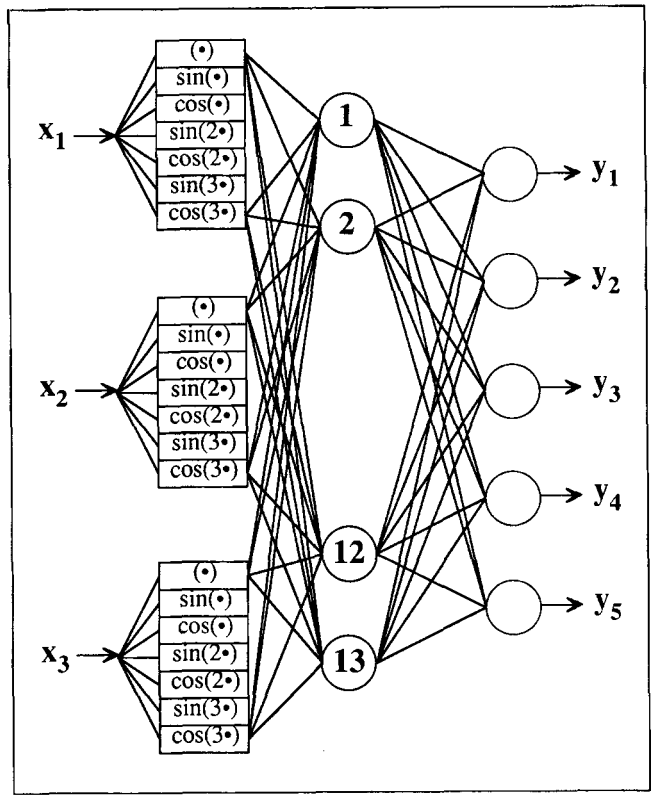

Figure 1. Configuration of modified network with functional input units.

It has been shown that multi-layered networks with nonlinear functions in neuron-like units can possess much better representation capabilities than two-layered networks (Lippmann, 1987). But such multi-layered networks also suffer, to some extent, limitations similar to those of two-layered networks. Our experience has indicated that the capability of the network to represent nonlinear relationships does not demonstrate notable improvements when the number of hidden layers is increased to more than two. This limitation is, perhaps, primarily due to the quasi-linearity of the sigmoidal logistic function $F(x-\theta)=1/(1+\exp[-x+\theta])$ in the neuron-like units (Lippman, 1987; Pao, 1989).

On the other hand, it can be realized from the above discussion that the representation capability of a network is primarily determined by the size of its input space and, hence, this capability may be enhanced by expanding the input space. This can be accomplished by adding to the input layer a number of functional units performing orthonormal transform actions of the inputs, such as $\sin(\gamma x)$, $\cos(\gamma x)$, $\sin(2\gamma x)$, $\cos(2\gamma x)$, and so on. Although this method does not require additional information from external measurements, it effectively increases the representation capability of a network. Even a two-layered network can be modified to represent nonlinear relationships of any complexity if a sufficient number of the functional units are used (Pao, 1989). The obvious disadvantage of this method is that network computation is increased. Thus, an appropriate compromise must be reached in the selection of the number of functional units and the number of hidden layers of a network. The architecture of the modified network used in this study is shown in Figure 1. The network consists of three layers: one output layer, one hidden layer and one

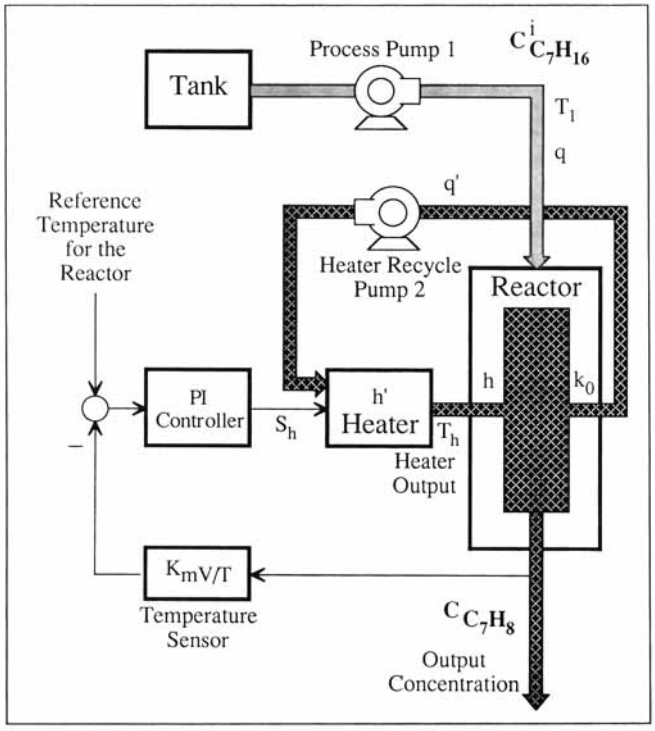

Figure 2. Process schematic.

input layer. The number of hidden nodes is also selected empirically, as the average between the number of total units and that of the output nodes. Each input is processed by sinusoidal functional units whose number is determined by trial and error through simulations.

Simulations of Fault Diagnosis in a Chemical Process

The process considered is a heptane-to-toluene aromatization system (Figure 2) near steady-state operation (Watanabe et al., 1989). The mathematical models describing the reactor, heater and controller are given in the Appendix. Detailed discussion about the process model can be found in Watanabe et al. (1989).

We assume that possible process faults are all at incipient stages. Control action prevents these faults from having an immediate impact on the reactor temperature. The severity of a fault is characterized by change(s) in corresponding parameter(s) in the process model. Five possible causes of faults are considered:

Fault 1: Deterioration of catalyst performance, leading to a decrease in the frequency factor for the catalyst, k_0, in the model.

Fault 2: Fouling of the heat exchanger surface in the reactor, leading to a decrease of overall heat-transfer coefficient, h, in the model.

Fault 3: Fouling of the heat exchanger surface in the heater, leading to a decrease of overall heat-transfer coefficient, h', in the model.

Fault 4: Partial plugging of the pipeline connected to pump 1, leading to a decrease in the volumetric flow rate, q, in the model.

Fault 5: Partial plugging of the pipeline connected to pump 2, leading to a decrease in the volumetric flow rate, q', in the model.

It is also assumed that only single and double faults are possible to occur during operation. Operation performance of the process is quantified through measurements of the outlet concentration of $C_{C_7H_8}$, the heater outlet temperature T_h and the controller output signal S_h. In practice, the relationships between measurements and fault causes may be obtained from historical and/or on-line data, with particular focus on records obtained when faults occurred. In this study, those relationships are obtained through numerical simulations. Single or double faults are simulated by decreasing the values of one or two of the fault-related parameters k_0, h, h', q, and q' in the model. All possible fault causes considered are examined for three different severity levels. A fault at level 1 is slight and

Table 1. Fault-to-Performance Relations of the Process (Subject to Single Fault)

Fault No./Level	Fault causes (Faulty parameters)*					Process Variable Changes		
	k_0^*	h^*	h'^*	q^*	q'^*	T_h^*	$C^*_{C_7H_8}$	S_h^*
normal	1.0	1.0	1.0	1.0	1.0	1.00	1.00	1.00
1/1	0.9	1.0	1.0	1.0	1.0	1.00	0.95	0.98
1/2	0.8	1.0	1.0	1.0	1.0	0.99	0.89	0.96
1/3	0.7	1.0	1.0	1.0	1.0	0.98	0.83	0.93
2/1	1.0	0.9	1.0	1.0	1.0	1.02	1.00	1.07
2/2	1.0	0.8	1.0	1.0	1.0	1.04	1.00	1.17
2/3	1.0	0.7	1.0	1.0	1.0	1.07	1.00	1.29
3/1	1.0	1.0	0.9	1.0	1.0	1.00	1.00	1.11
3/2	1.0	1.0	0.8	1.0	1.0	1.00	1.00	1.25
3/3	1.0	1.0	0.7	1.0	1.0	1.00	1.00	1.43
4/1	1.0	1.0	1.0	0.9	1.0	0.99	1.05	0.95
4/2	1.0	1.0	1.0	0.8	1.0	0.97	1.11	0.90
4/3	1.0	1.0	1.0	0.7	1.0	0.96	1.17	0.84
5/1	1.0	1.0	1.0	1.0	0.9	1.00	1.00	0.90
5/2	1.0	1.0	1.0	1.0	0.8	1.00	1.00	0.80
5/3	1.0	1.0	1.0	1.0	0.7	1.00	1.00	0.70

* The superscript * denotes that the value of a variable or a parameter is normalized on the basis of its value at normal condition.

corresponds to 10 percent decrease of the fault-related parameter from its normal value. A fault at level 2, corresponding to 20 percent decrease, is medium and a fault at level 3, corresponding to 30 percent decrease, is severe. Table 1 and Figure 3 show the simulation results of fault-to-performance relations for the single fault case. Figure 3a identifies fault causes with values of possible faulty parameters, and Figure 3b shows the effect of these faulty parameters on process variables. Figure 4 shows double-fault simulation results.

A neural network with modified architecture through the addition of functional units (Figure 1) is utilized for fault detection. The network has five output nodes, corresponding to the five possible faulty parameters mentioned above. The three measured process variables correspond to three network inputs. Each one of these three measurements is channelled to seven input nodes, six of which are the added functional units. That number was selected by trial and error. The number of hidden nodes was empirically chosen to be the average between the number of input ($3 \times 7 = 21$) and output (five) nodes. The input variables are normalized as

$$\bar{x}_i = 5\frac{x_i - x_i^0}{x_i^0} \tag{7}$$

with $x_1 = T_h$, $x_2 = C_{C_7H_8}$, $x_3 = S_h$, and x_i^0 denoting normal operation values of corresponding variables. The scaling factor 5 is selected such that the normalized variables $\bar{x}_i$ can fully map the region $(-3, 3) \approx (-\pi, \pi)$, a desired input range for the network.

The fault-to-performance relations shown in Figures 3 and 4 are utilized to generate training patterns for the network, with the normalized variables, $\bar{x}_i$, defined by Eq. 7 as inputs. The outputs of the network, standing for individual fault types, take values encoding different severity levels of the corresponding faults, as shown in Table 2, where the value of 0.2 is assigned to stand for normal condition, 0.7 for a fault at level 1, 0.8 for a fault at level 2, and 0.9 for a fault at level 3.

The generated training patterns are repeatedly presented on the network, and the network connection weights, w_{ij}, together with the activation thresholds, θ_j, are iteratively corrected until the criterion

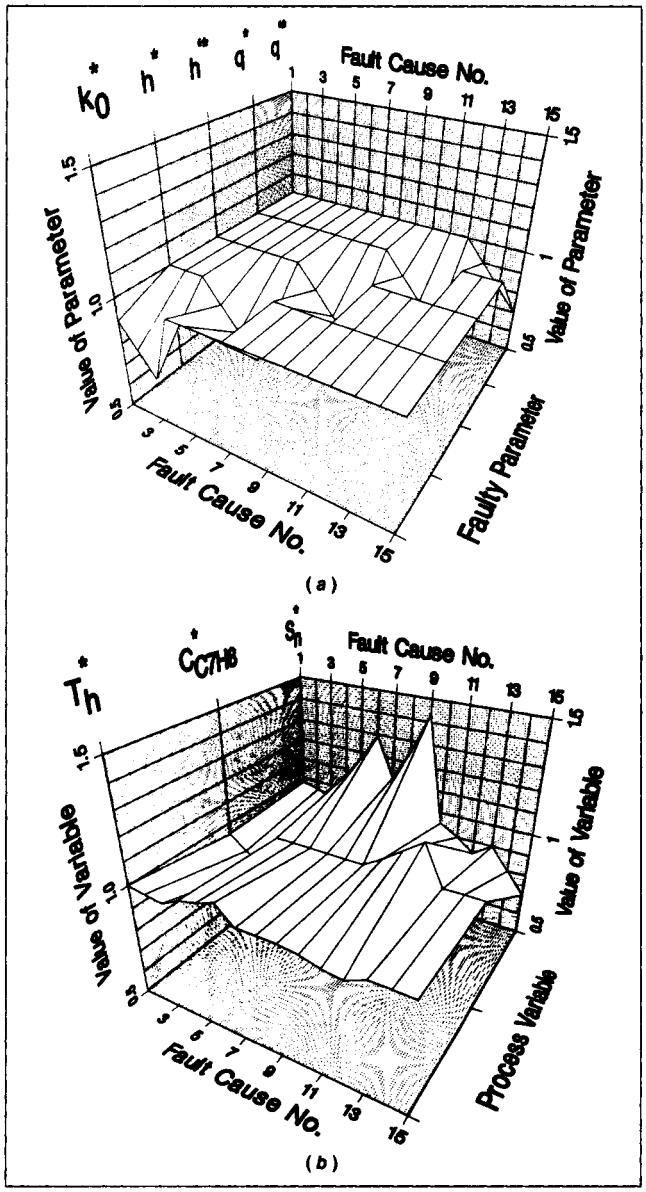

Figure 3a,b. Fault-to-performance relations of the process (subject to single fault).

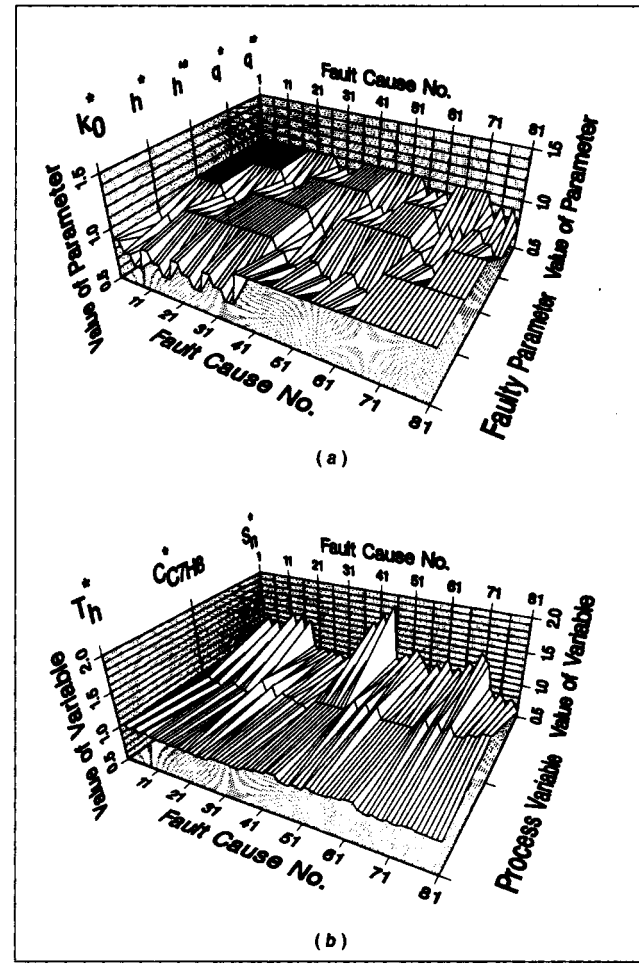

Figure 4a,b. Fault-to-performance relations of the process (subject to double fault).

Table 2. Encoding Values for Fault Levels

Normal	Level 1	Level 2	Level 3
0.20	0.70	0.80	0.90

$$|t_k^p - y_k^p| < 0.01 \qquad (8)$$

is met for the individual outputs under all the training patterns. Here t represents the desired output as specified in Table 2, y represents the calculated output of the network, the superscript p stands for each training pattern, and the subscript k for each output of the network in its output layer. The learning rate and the momentum constant are specified as $\eta = 0.2$ and $\alpha = 0.5$, respectively. The quantity β of activation functions in the neuron-like units is specified as $\beta = 1$. The network connection weights, w_{ij}, and the thresholds θ_j, are initialized randomly with values in the range $[-1.5, 1.5]$. A bias node of unity is employed for the hidden layer and the output layer so that the thresholds can be updated for adjusting the weights to that node as in other nodes. It takes approximately 12,000 iterations before the criterion of Eq. 8 is met. Each training pattern is presented once at every iteration.

Table 3 and Figure 5 show the trained network activations for part of the faulty operation case (single faults) considered in the training patterns. It is observed from Table 3 that the network gives correct detections for the fault causes considered, and precisely realizes the corresponding fault levels simultaneously, a task that the conventional network by Watanabe et al. (1989) exhibited an inability to perform. Similar behavior is observed in the double-fault case (Figure 6).

A set of novel fault patterns, different from the training ones, with fault severities at levels 1½, 2½, and 3½ were also tested on the network. Table 4 and Figure 7 illustrate part of the test results for the single-fault case, while Figure 8 shows the results for the double-fault case. Correct detections were again observed. It is also observed that the results at level 3½ are not as good as those at other levels. This is a common problem of networks when doing extrapolation outside the training region. However, in the application under consideration, it may not cause a significant inconvenience because we have already defined level 3 to be the most severe fault to which, once it occurs, corresponding emergency measures have to be taken. A fault more severe than level 3 will be handled in the same way. Thus, it is not necessary to require the network to do further discrimination.

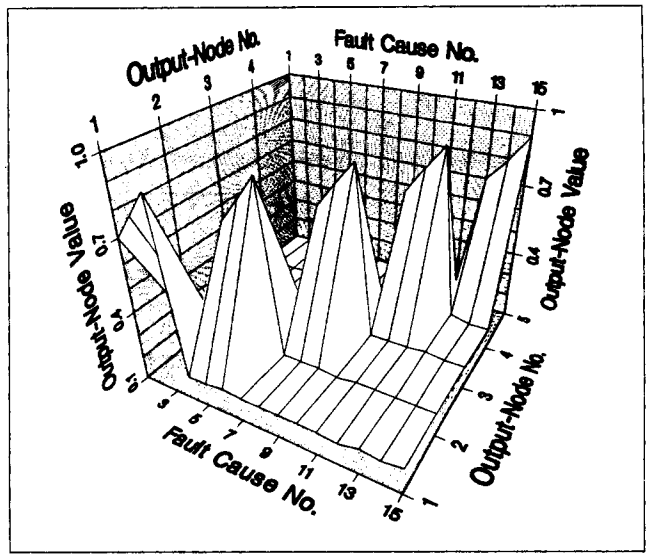

Figure 5. Diagnosis results of the network (single-fault cases as in training phase).

The successful results shown above imply that the modified network indeed effectively expands its nonlinear representation capability. It can be used to perform general detection for possible faults within consideration once it is trained properly by appropriate representative fault operating patterns.

In practical applications, measurement noise will inevitably exist. This may cause the output activation values of the network to be greater than 0.2 at normal condition. Thus, a threshold value, $\sigma = 0.6$, is employed to filter out the effects of measurement noise as follows: if an output activation value is less than σ, it will be considered to be normal. Simulations

Table 3. Diagnosis Results of the Network (Single-Fault Cases as in Training Phase)

Fault No./Level	Network Output Values on Output Nodes				
	Node 1	Node 2	Node 3	Node 4	Node 5
normal	0.19	0.21	0.20	0.19	0.20
1/1	0.70	0.19	0.20	0.20	0.20
1/2	0.80	0.20	0.20	0.21	0.20
1/3	0.90	0.20	0.20	0.20	0.20
2/1	0.20	0.70	0.20	0.21	0.20
2/2	0.20	0.80	0.20	0.20	0.20
2/3	0.21	0.90	0.21	0.20	0.20
3/1	0.20	0.20	0.70	0.20	0.19
3/2	0.20	0.20	0.80	0.19	0.19
3/3	0.20	0.20	0.90	0.19	0.20
4/1	0.20	0.19	0.20	0.70	0.20
4/2	0.20	0.20	0.20	0.81	0.20
4/3	0.19	0.20	0.20	0.91	0.20
5/1	0.21	0.20	0.21	0.21	0.70
5/2	0.20	0.20	0.20	0.19	0.79
5/3	0.21	0.20	0.20	0.19	0.90

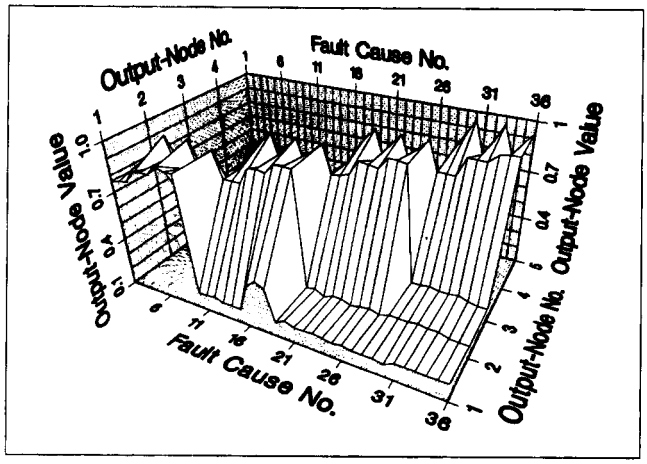

Figure 6. Diagnosis results of the network (double-fault cases as in training phase).

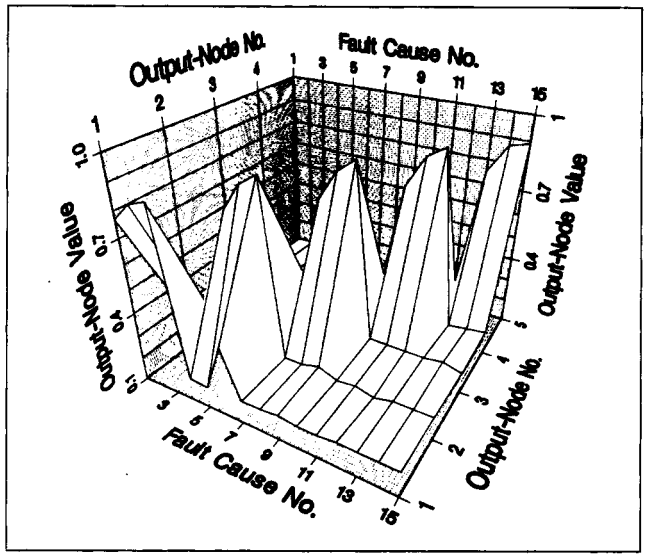

Figure 7. Diagnosis results of the network (single-fault cases in test phase).

under noisy conditions show that this criterion can effectively filter out the effects of measurement noise with standard deviation less than 3%.

Conclusion

An improved approach to fault diagnosis of chemical processes via neural networks was presented. The proposed technique, relying on modification of the conventional backpropagation network by adding functional units to the input layer, can significantly expand the nonlinear representation capability of a network. This improvement makes it possible to perform simultaneous detection of multiple fault causes and their severity levels for chemical processes. Applicability of the proposed approach to practical problems was illustrated by simulations of successful fault detection for a selected example process.

Notation

a = area of heat exchange, 10 m^2
$C_{C_7H_8}$ = outlet concentration of C_7H_8, 524 gmol/m^3
$C_{C_7H_{16}}$ = outlet concentration of C_7H_{16}, 476 gmol/m^3

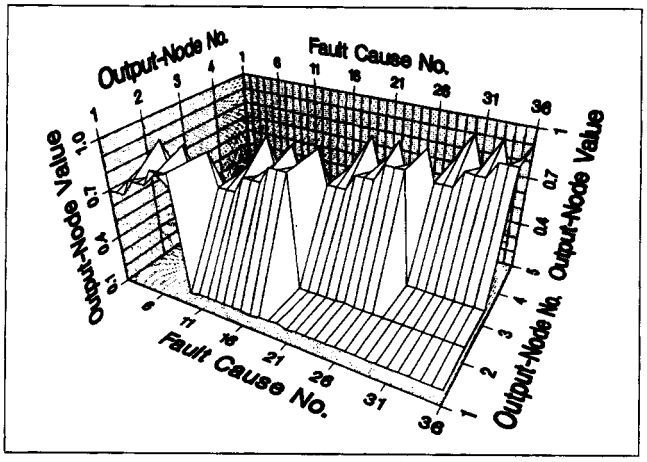

Figure 8. Diagnosis results of the network (double-fault cases in test phase).

Table 4. Diagnosis Results of the Network (Single-Fault Cases as in Test Phase)

Fault No./Level	Network Output Values on Output Nodes				
	Node 1	Node 2	Node 3	Node 4	Node 5
1/1½	0.77	0.19	0.20	0.20	0.20
1/2½	0.84	0.20	0.20	0.21	0.20
1/3½	0.87	0.20	0.19	0.18	0.24
2/1½	0.20	0.75	0.20	0.20	0.20
2/2½	0.19	0.86	0.20	0.20	0.20
2/3½	0.40	0.91	0.39	0.21	0.21
3/1½	0.20	0.20	0.74	0.20	0.19
3/2½	0.20	0.20	0.86	0.19	0.19
3/3½	0.21	0.22	0.92	0.33	0.19
4/1½	0.20	0.19	0.20	0.74	0.20
4/2½	0.20	0.20	0.20	0.87	0.21
4/3½	0.21	0.18	0.20	0.92	0.22
5/1½	0.21	0.20	0.21	0.18	0.75
5/2½	0.21	0.20	0.21	0.19	0.87
5/3½	0.21	0.20	0.19	0.20	0.89

$C^i_{C_7H_{16}}$ = inlet concentration of C_7H_{16}, 1,000 gmol/m^3
C_p = specific heat, 490.7 J/gmol K
E_a = activation energy, 1.369×10^5, J/gmol
h = overall heat-transfer coefficient, 6.05×10^5 J/m^2 h K
ΔH = heat of reaction, $=2.2026\times10^5+6.2044\times10^1T-5.536\times10^{-2}T^2-1.15\times10^{-6}T^3+3.1496\times10^{-7}T^4$, J/gmol
k_c = proportional gain of controller, 20
K = $a'h'k'_h/(\rho' C'_p q')$, 1.0 K/mV
$K_{mV/T}$ = gain of temperature to mV transducer, 1.0 mV/K
k_0 = frequency factor, 5.01×10^8, h^{-1}
net_i = input of a neuron-like unit in the network
q = inlet and outlet volumetric flow rate, 3 m^3/h
R = gas constant, 8.319 J/gmol K
S_h = output of the PI controller, 223 mV
S_i = output of integrator in PI controller, 223 mV
T = reaction temperature, 740 K
T_c = inlet temperature of heating water, 0.9T K
T_h = heater outlet temperature, 889 K
T_i^* = integrator coefficient, 0.3 h
T_i = temperature of reaction inlet stream, 300 K
u_c = command signal of controller, 740 mV
V = effective reactor volume, 30 m^3
w_{ij} = network connection weight between two units
x_i = signal from external system to the network input layer
$\bar{x}_i$ = dimensionless input variable defined in Eq. 7
y_i = output of a neuron-like unit in the network

Greek letters

α = momentum constant
β = transition abruptness of the activation function
η = learning rate
θ_j = threshold of an activation function
ρ = density, 593 gmol/m^3
σ = threshold for eliminating noise effect
τ = heater time constant, 0.2 h

Literature Cited

Hoskins, J. C., and D. M. Himmelblau, "Artificial Neural Network Models of Knowledge Representation in Chemical Engineering," *Computers and Chem. Eng.*, **12**, 881 (1988).

Himmelblau, D. M., "Fault Detection and Diagnosis in Chemical and Petrochemical Processes," *Chem. Eng. Monographs*, No. 8, Elsevier (1978).

Lippmann, R. P., "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine*, **4**(2), 4 (April 1987).

McLelland, J. L., and D. E. Rumelhart, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, MIT Press, Cambridge, MA (1986).

Pao, Yoh-Han, *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley (1989).

Venkatasubramanian, V., and K. Chan, "A Neural Network Methodology for Process Fault Diagnosis," *AIChE J.*, **35**, 1993 (1989).

Watanabe, K., I. Matsuura, M. Abe, M. Kubota, and D. M. Himmelblau, "Incipient Fault Diagnosis of Chemical Processes via Artificial Neural Networks," *AIChE J.*, **35**, 1803 (1989).

Appendix

The process shown in Figure 2 is mathematically modeled as follows:

$$\frac{dT}{dt}=\frac{q}{v}(T_i-T)-\frac{\Delta H}{\rho C_p}k(T)\,C_{C_7H_{16}}+\frac{ah}{\rho C_p v}(T_h-T)$$

$$\frac{dC_{C_7H_8}}{dt}=-\frac{q}{v}C_{C_7H_8}+k(T)\,C_{C_7H_{16}}$$

$$\frac{dC_{C_7H_{16}}}{dt}=-\frac{q}{v}C_{C_7H_{16}}-k(T)\,C_{C_7H_{16}}+\frac{q}{v}C^i_{C_7H_{16}}$$

$$k(T)=k_0\exp\left[\frac{E_a}{RT}\right]$$

$$\frac{dT_h}{dt}=\frac{1}{\tau}(T_c-T_h)+\frac{K}{\tau}S_h$$

$$\frac{dS_h}{dt}=k_c(u_c-K_{mV/T}T)+S_i$$

$$\frac{dS_i}{dt}=\frac{k_c}{T_i^*}(u_c-K_{mV/T}T)$$

Numerical values are given in the notation section.

Manuscript received July 12, 1990, and revision received July 20, 1992.