ASHRAE Transactions V.102, Pt. 1, pp. 540 - 549, 1996

AT-96-3-3

Fault Diagnosis of an Air-Handling Unit Using Artificial Neural Networks

Won-Yong Lee, Ph.D.

John M. House, Ph.D.
Associate Member ASHRAE

Cheol Park, Ph.D.

George E. Kelly, Ph.D.
Fellow ASHRAE

ABSTRACT

The objective of this study is to describe the application of artificial neural networks to the problem of fault diagnosis in an air-handling unit. Initially, residuals of system variables that can be used to quantify the dominant symptoms of fault modes of operation are selected. Idealized steady-state patterns of the residuals are then defined for each fault mode of operation. The steady-state relationship between the dominant symptoms and the faults is learned by an artificial neural network using the backpropagation algorithm. The trained neural network is applied to experimental data for various faults and successfully identifies each fault.

INTRODUCTION

Modern buildings are being designed with increasingly complex operating systems that have seemingly limitless capabilities for monitoring and controlling the conditions in the building. Unfortunately, building operators are not always able to monitor and process the enormous amounts of data that are generated. Hence, there is a need for robust fault detection and diagnostic tools that can be used to assist the building operator and ensure that the system is operating in the manner in which it was designed. The benefits of a properly operating building system are numerous, including improved energy efficiency, improved occupant comfort and health, and longer equipment life.

In a companion paper, Lee et al. (1996) describe methods for fault detection in an air-handling unit (AHU). One approach used in that study is to define residuals that represent the difference between the existing state of the system and the normal state. Residuals that are significantly different from zero represent the occurrence of a fault. If the system that is being monitored is not too complex, the building operator should have little trouble isolating the source of the fault after the initial detection. However, for complex systems, isolating the fault can be challenging and diagnostic tools are needed. This paper describes the use of artificial neural networks (ANNs) for this purpose.

Several studies that examine the use of ANNs for fault diagnosis appear in the literature. Watanabe et al. (1994) and Fan et al. (1993) used ANNs for fault diagnosis of chemical processes. Watanabe et al. (1994) proposed a two-stage, multilayer ANN. In the first stage, the faults were diagnosed, and in the second stage, the degree of the fault was estimated. The study of Fan et al. (1993) was based on steady-state operating conditions. Koive (1994) reviewed studies that utilized ANNs for fault diagnosis and control and summarized the architectures most widely used in practice. The paper also summarized steady-state and dynamic fault diagnosis and control for a paper-making machine.

Fault diagnosis can be thought of as pattern recognition, and ANNs are well suited to this task. For example, ANNs using the backpropagation algorithm can be used for character recognition (Ananthraman 1995; Demuth and Beale 1992). The input patterns are matrix representations of the dark (1's) and light (0's) pixels of the 26 characters of the alphabet, and the output patterns are 26-bit strings of 1's and 0's that represent the various characters. A similar approach can be applied for fault diagnosis. Normal and fault modes of operation typically have operational signatures or distinguishing patterns for each mode of operation. ANNs can learn and exploit these patterns to diagnose the current operational mode of a system. The objective of this study is to describe the application of an ANN to the problem of fault diagnosis in an AHU. As an intermediate step, a second ANN is used as a process model for a cooling coil valve subsystem.

The first two sections of the paper provide a brief description of the AHU and the residuals used in the fault diagnosis. The eight faults and their corresponding symptoms and dominant residuals are then described. Next, a brief description of ANNs and the backpropagation algorithm is provided. Applications of ANNs to the development of a model of the cooling coil valve subsystem and fault diagnosis are then discussed. Finally, results

Won Yong Lee is a senior researcher at the Korea Institute of Energy Research, Taejon, South Korea. **John M. House** and **Cheol Park** are mechanical engineers and **George E. Kelly** is a group leader in the Mechanical Systems and Controls Group, Building Environment Division, Building and Fire Research Laboratory, National Institute of Standards and Technology, Gaithersburg, Md.

of the fault diagnosis are presented and conclusions and recommendations for future work are discussed.

AIR-HANDLING UNIT

A schematic diagram of the variable-air-volume (VAV) AHU utilized for this study is shown in Figure 1. The same system was used for a companion paper on fault detection (Lee et al. 1996). The AHU consists of fans, dampers, a cooling coil, sensors, and controllers. The static pressure in the main supply duct is maintained at a constant setpoint value of 249 Pa (1.0 in. H_2O) by sensing the static pressure and controlling the rotational speed of the supply fan. The supply air temperature is controlled by modulating the cooling-water control valve to maintain a constant setpoint value of 14.5°C (58.0°F). The airflow rate difference between the supply and return airstreams is controlled by the variable-speed return fan to maintain a constant setpoint value of 0.472 m^3/s (1000.0 cfm). A proportional-integral-derivative (PID) algorithm is used to control the cooling water valve, and PI algorithms are used to control the supply and return fan speeds. Although not shown in Figure 1, a personal computer (PC) and a data-acquisition system (DAS) are used for purposes of computing control signals and logging data. The sampling period for control and data collection is 10 seconds.

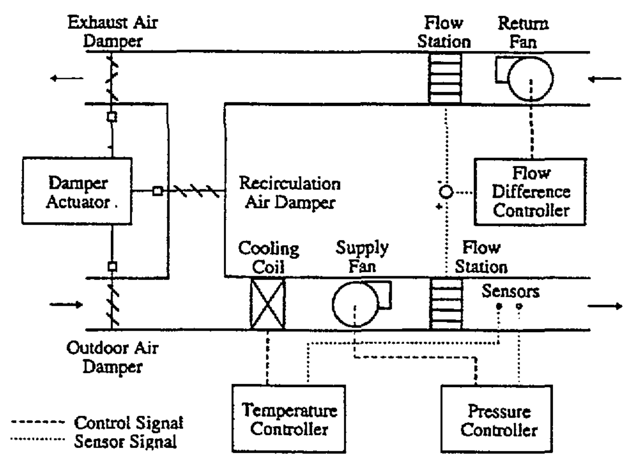

Figure 1 Schematic diagram of a variable-air-volume air-handling unit.

RESIDUAL DEFINITION

The approach used in this paper relies on the ability to identify patterns of residuals that can be used as signatures for various faults. Through laboratory testing, it was determined that seven residuals are needed to identify the eight faults considered here (described in the next section). The first three residuals represent the difference between actual and setpoint values of the supply air temperature, supply air pressure, and the airflow rate difference between the supply and return ducts. The residuals are given by

$$R_T = T_S - T_{S,SP}, \quad (1)$$

$$R_P = P_S - P_{S,SP}, \quad (2)$$

$$R_Q = Q_D - Q_{D,SP} \quad (3)$$

where

R = residual value,
T_S = supply air temperature,
P_S = supply air pressure,
Q_D = airflow rate difference in the supply and return ducts,
$T_{S,SP}$ = setpoint value of T_S,
$P_{S,SP}$ = setpoint value of P_S, and
$Q_{D,SP}$ = setpoint value of Q_D.

The cooling coil valve control signal can provide valuable insight into the operating status of the AHU. A residual is defined for the operation of the cooling coil valve and is given by

$$R_U = U_{CC} - U_{CC,EV} \quad (4)$$

where

U_{CC} = actual control signal to the cooling coil valve,
$U_{CC,EV}$ = expected value of U_{CC}.

U_{CC} is determined by the PID controller for the supply air temperature; however, there is no obvious way to specify $U_{CC,EV}$. In this study $U_{CC,EV}$ is determined using an ANN model of the cooling coil valve subsystem. The model for $U_{CC,EV}$ is described in a later section.

Residuals for the operation of the supply and return fans are given by

$$R_{NS} = N_S - U_S \quad (5)$$

$$R_{NR} = N_R - U_R \quad (6)$$

where

N_S = measured value of the supply fan speed,
N_R = measured value of the return fan speed,
U_S = control signal for supply fan, and
U_R = control signal for return fan.

The final residual is based on a comparison of the actual cooling coil valve position and the expected value based on the actual cooling coil control valve signal. The residual is given by

$$R_V = V_p - U_{CC} \quad (7)$$

where

V_p = two-way cooling coil valve position.

Residuals such as R_U require the comparison of measured values to model values and can cause difficulties arising from the use of models. The most obvious problem is model error. Even if the model is accepted as being an accurate representation of the physical process, it may require the identification of parameter values specific to each physical system. In addition, the characteristics of the system can change over time and require that the parameters be identified periodically. However, these models and the residuals based upon them contain the underlying physics of the process(es) involved. They provide information on state variables, such as the rate of heat transfer from a coil, that cannot be (easily) measured directly. Thus, it is not practical to think that such model-based residuals can be eliminated. As stated previously, and as will become more apparent in the next section, R_U is a symptom of several of the faults studied here. Hence a reliable model of the operation of the cooling coil valve under normal operating conditions is needed, and part of this study is devoted to a discussion of the use of an ANN for this purpose.

FAULT DESCRIPTION

Faults are typically classified as belonging to one of two categories, namely, faults due to a complete failure of a component or system and faults due to performance degradation. The main factor used in categorizing faults is the rate at which they occur. Complete failures typically occur abruptly, although they may be due to factors such as equipment wear that take place over years of use. Faults involving performance degradation evolve over periods that are typically measured in weeks, months, or years. Faults of this nature are difficult to detect in their early stages because only subtle changes occur in the component or system performance. The faults considered in this study represent complete failures of components of the system. Complete failures are considered because they can be easily introduced in a laboratory system and the fault symptoms can be observed almost immediately. Simulation studies are more appropriate for faults caused by performance degradation.

Eight faults representing complete failures of various components in the AHU are described. The dominant symptoms of each fault are also described. The faults are introduced when the system is operating at normal, steady-state conditions, and the dominant symptoms correspond to the steady-state conditions after a fault has occurred. With the exception of the pump fault, all faults are simulated in the laboratory AHU by either sending faulty control signals from the PC to an actuator or by overwriting sensor signals that are logged by the DAS with faulty values. In each case, the faulty control or sensor signal is equal to its minimum possible value (usually zero). The pump fault is introduced manually by reducing the pressure of the cooling water supplied to the cooling coil.

Fault 1 is a failure of the supply fan. During normal operation the supply fan is controlled to maintain a static pressure of 249 Pa (1.0 in. H_2O) in the supply air duct, and the return fan is controlled to maintain a flow difference of 0.472 m^3/s (1,000 cfm) between the supply and return air ducts. The fault causes the supply fan rotational speed to decrease to zero, the supply air pressure to decrease to zero, and the control signal to the supply fan to increase to its maximum value in an attempt to offset the decreasing supply air pressure. The control signal for the return fan decreases to zero in an attempt to maintain the flow difference between the supply and return air ducts at the setpoint temperature; however, this condition cannot be achieved due to the fault. Because there is no airflow, the supply air temperature gradually increases, resulting in the cooling coil valve control signal increasing to its maximum value. Thus, the dominant residuals for fault 1 are R_P, R_Q, R_U, and R_{NS}.

Fault 2 is a failure of the return fan. The fault causes the return fan rotational speed to decrease to zero, the flow difference between the supply and return ducts to increase, and the control signal to the return fan to increase to its maximum value in an attempt to offset the increasing flow difference. Thus, the dominant residuals for fault 2 are R_Q and R_{NR}.

Fault 3 is a failure of a chilled-water pump. It is assumed that more than one pump is used to deliver the chilled water to the AHU and therefore the fault causes the water flow rate to decrease, but not to zero. The decrease in the flow rate of cooling water causes the supply air temperature to increase initially. This causes the cooling coil valve signal to increase, thus opening the valve. By opening the cooling coil valve it may be possible to bring the supply air temperature back to the setpoint value; however, the control signal to the cooling coil valve will be different from the normal condition. The dominant residual for fault 3 is R_U.

Fault 4 is a stuck cooling coil valve. After the fault is introduced, nearly normal operation continues and the residuals remain near zero until a disturbance occurs that calls for a significant change in the valve position. As an example, a load decrease in a zone causes the damper of the VAV box for that zone to close, thus increasing the static pressure in the supply duct. This causes the supply fan speed to decrease to bring the static pressure back down to the setpoint value and, consequently, the supply air temperature decreases. In an attempt to compensate for the decreasing supply air temperature, a signal is sent to the cooling coil valve to close further. However, because the valve is stuck, the position does not change. Over time, the integrator portion of the control algorithm causes the cooling coil control signal to decrease to its minimum value. The dominant residual for fault 4 is R_V.

Fault 5 is a failure of the supply air temperature thermocouple. A thermocouple failure typically results in a voltage signal that varies randomly between large positive and negative values. If the sensed value of the supply air temperature is outside the range of normal operating conditions (0°C to 40°C, for example), the temperature could be automatically set to zero so that the residual R_T would not fluctuate. This type of failure is simulated by overwriting the sensed supply air temperature with a value of 0°C. A zero supply air temperature signal causes the signal to the cooling coil control valve to decrease to its minimum value and thus close the valve in an attempt to raise the supply air temperature. The dominant residuals for fault 5 are R_T and R_U.

Fault 6 is a failure of the supply air pressure transducer. When this failure occurs, a zero reading is obtained for the supply air pressure (by overwriting the actual value). This causes the control signal to the supply fan to increase to its maximum value in an attempt to increase the supply pressure. The supply airflow rate increases for a short period (until the VAV boxes respond) and this makes it necessary for the cooling coil valve to open further to maintain the supply air temperature at the setpoint value. The return fan control signal also increases for a short period in order to maintain the flow rate difference between the supply and return ducts at the setpoint value. The dominant residual for fault 6 is R_P.

Fault 7 is a failure of the supply fan flow station. When this fault occurs, a zero reading is obtained for the supply flow station (by overwriting the actual value) and the return fan controller believes that there is no flow in the supply duct. Thus, the control signal to the return fan is decreased to its minimum value in an attempt to maintain the flow difference between the supply and return ducts at the setpoint value. However, the measured flow difference can only approach zero and does not reach the setpoint value because this would require a negative flow of 0.472 m^3/s (1,000 cfm) in the return duct. The dominant residual for fault 7 is R_Q.

Fault 8 is a failure of the return fan flow station. When this fault occurs, a zero reading is obtained for the return flow station (by overwriting the actual value) and the return fan controller believes that there is no flow in the return duct. Thus, the control signal to the return fan is increased to its maximum value in an attempt to maintain the flow difference between the supply and return ducts at the setpoint value. However, the measured flow rate difference is unchanged by this compensation due to the presence of the fault. The dominant residual for fault 8 is R_Q.

ARTIFICIAL NEURAL NETWORKS

Introduction

The ANNs used in this study have a multilayer feedforward network structure and are trained using a backpropagation learning rule. Multilayer feedforward networks consist of an input layer, an output layer, and one or more hidden layers. A schematic diagram of a multilayer feedforward network with one hidden layer is shown in Figure 2. The inputs to the n_i input units are denoted $x_1, x_2, \dots x_{n_i}$; the outputs of the n_o output units are denoted $y_1, y_2, \dots y_{n_o}$; and outputs of the n_h hidden layer units are denoted $h_1, h_2, \dots h_{n_h}$. The nonshaded units are bias units whose inputs are set equal to unity. The connections between the units of different layers of the network are weights and biases. The variable names assigned to particular weights and biases are given in Figure 2 and correspond to the dotted line connections in the figure. The ANN is trained to learn the functional mapping of inputs to outputs using input/output training pairs. The output training data are referred to as the target output of the ANN. The goal is to train the network until the output of the ANN is suitably close to the target output (Hertz et al. 1991).

Consider initially the forward pass through the network. For a specific input pattern (set of input values), the output of the jth hidden layer unit is given by

$$h_j = f\left(\sum_{i=1}^{n_i} w'_{ji} x_i + b'_j\right) \tag{8}$$

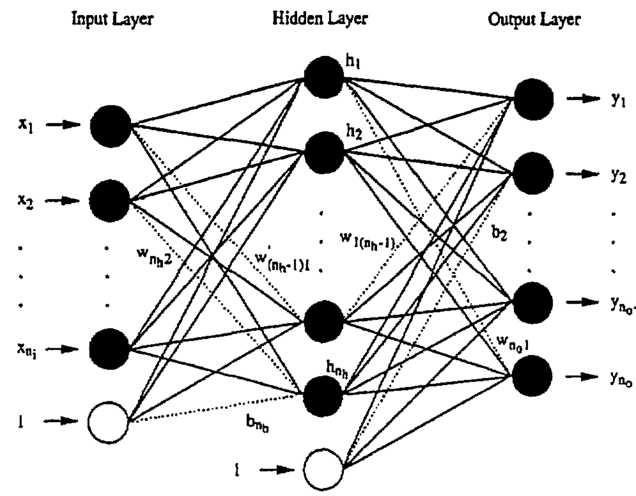

Figure 2 Two-layer feedforward network.

where

f = activation function,

w'_{ji} = strength of connection from ith input unit to jth hidden layer unit,

b'_j = bias value for jth hidden layer unit.

The output of the kth output unit is given by

$$y_k = f\left(\sum_{j=1}^{n_h} w_{kj} h_j + b_k\right) \tag{9}$$

where

w_{kj} = strength of connection from jth hidden layer unit to kth output unit and

b_k = bias value for kth output unit.

The backpropagation algorithm uses a gradient descent algorithm to update the weights, and therefore the activation functions must be differentiable. The activation functions used for the ANNs in this study are

$$f(x) = x \tag{10a}$$

$$f(x) = \frac{1}{1 + e^{-x}} \tag{10b}$$

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{10c}$$

where the functions given by Equations 10a through 10c are referred to as the pure linear function, the log-sigmoid function, and the tan-sigmoid function, respectively. The result of the forward pass is the output pattern $y_1, y_2, \dots y_{n_o}$.

As stated previously, training is continued until the output patterns are suitably close to the target patterns. Mathematically this is achieved by minimizing the sum-of-squares error (SSE), given by

$$SSE = \sum_{p=1}^{n_p} \sum_{k=1}^{n_o} (t_{k,p} - y_{k,p})^2 \tag{11}$$

where

$t_{k,p}$ = target value for the kth output unit of the pth pattern,

$y_{k,p}$ = actual value for the kth output unit of the pth pattern, and

n_p = total number of training patterns.

From Equation 11, SSE is computed by summing over all n_o output values for all n_p training patterns.

The ANN is trained by updating the weights using a backpropagation learning rule. The change in weight (ω'_{ji}) is based on the gradient descent rule and is given by

$$\Delta w'_{ji} = -\eta \frac{\partial(SSE)}{\partial w'_{ji}} \tag{12}$$

where

η = learning rate.

A more complete description of ANNs and the backpropagation algorithm is given by Hertz et al. (1991).

Application of an ANN to the Cooling Coil Valve Subsystem

To compute residual R_U, a model is needed to determine the expected value of the cooling coil valve control signal $U_{CC,EV}$. An ANN can also be utilized for this purpose. Curtiss et al. (1993) described the modeling of a heating coil using a neural network where the objective was to determine the load on the coil for the next time step. For this study the goal is to determine the current value of U_{CC} for normal operating conditions. A schematic diagram of the cooling coil and the cooling coil valve subsystem is shown in Figure 3. T_M and ϕ_M are the mixed air temperature and relative humidity, respectively; Q_S is the supply airflow rate; and T_{WI} is the temperature of the cooling water at the inlet to the cooling coil. The other variables retain their previous definitions.

The ANN used to model the cooling coil valve subsystem has a single hidden layer with 10 units. Knowledge of the physical process and extensive training and testing of different combinations of input variables and network topologies were utilized to identify the inputs to the ANN. As discussed later in this section, the inputs represent a tradeoff between performance of the ANN under normal and faulty conditions. The input and output variables for training are:

Inputs $Q_S(i)$, $Q_S(i-1)$, $T_S(i)$, $T_S(i-1)$, $T_M(i)$, $T_M(i-1)$, $T_{WI}(i)$, $T_{WI}(i-1)$, $\phi_M(i)$, $\phi_M(i-1)$, $Q_S(i)\,[T_M(i) - T_S(i)]$, $Q_S(i-1)\,[T_M(i-1) - T_S(i-1)]$

Output $U_{CC}(i)$

where (i) refers to the current discrete time value and $(i-1)$ refers to the previous value. Inputs of the form $Q_S(i)\,[T_M(i) - T_S(i)]$ are measures of the load on the coil at a particular time. The ANN is trained in a batch mode (off-line) using experimental data obtained as the system operates in a normal mode. The training data consist of 2,271 input/output patterns, and training proceeds until the average error for each training pattern is approximately 0.0015. A tan-sigmoid activation function is used for the hidden layer and a pure linear activation function is used for the output layer. A commercial ANN software package is used for the training (Demuth and Beale 1992).

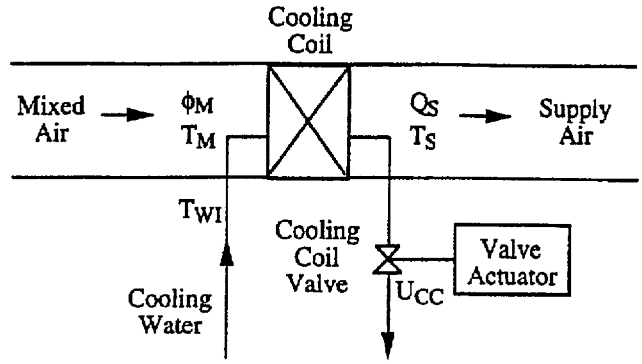

Figure 3 Schematic diagram of the cooling coil and cooling coil valve subsystem.

The actual value of U_{CC} and the ANN model value of $U_{CC,EV}$ are plotted as a function of time (denoted t) in Figure 4 for a stuck valve fault (fault 4). The fault and a load decrease are at t = 1,800 seconds. The load decrease at 1,800 seconds causes U_{CC} to decrease to its minimum value (1 V in this case) in an attempt to bring the supply air temperature up to the setpoint temperature. $U_{CC,EV}$ also shows a decrease at t = 1,800 seconds; this is due to the decrease in the supply airflow rate that occurs when the load decreases. A distinct difference in the two signals is observed and this difference is used to compute R_U. This plot demonstrates that the ANN model responds to normal system changes in the appropriate manner; however, the model does not respond to the changes caused by the fault.

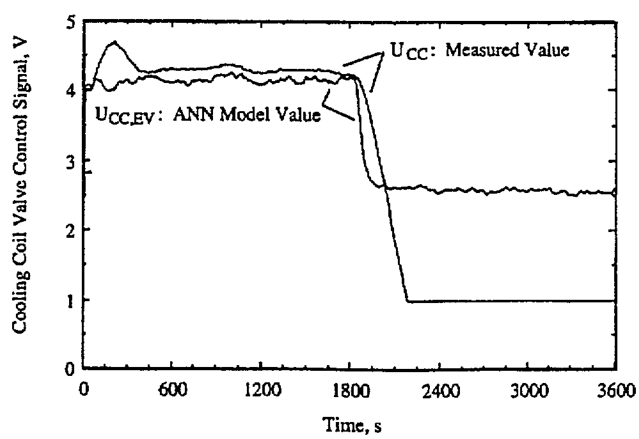

Figure 4 Actual and predicted cooling coil valve control signals.

For all testing, the supply air temperature inputs to the ANN are replaced by the supply air temperature setpoint. This is done to avoid contaminating the ANN inputs with faulty data. In addition, $Q_S(i)$ is monitored so that if its value goes to zero, the input to the ANN model is modified so that $Q_S(i)$ is equal to its average value from the previous 20 time steps. Hence, faulty data associated with the supply fan fault (fault 1) and the supply fan flow station fault (fault 7) have only a minimal effect on the computation of the expected value of the cooling coil valve control signal, $U_{CC,EV}$.

In general, the ANN model is susceptible to faulty input data, as would any model that uses real data as input. This is a key issue in the development of the ANN model because the goal is to predict the operation of the valve for normal conditions, not for fault conditions. For the latter case, a sufficiently trained ANN would simply track the faulty control signal and the residual R_U would not indicate the presence of a fault. In this study this was avoided by not using past values of U_{CC} as inputs to the ANN model.

The input training data do not exhibit a great deal of variation for T_{WI}, ϕ_M, and T_M and, therefore, the ANN can only be used reliably for a relatively small range of these variables. Future effort will be devoted to collecting data over a wider range of conditions; however, the current set of training data is sufficient to demonstrate this application of ANNs. The need for a large training data set that covers the complete range of operating conditions for the process is a practical consideration that must be overcome to implement this model.

Application of an ANN to Fault Diagnosis

To utilize an ANN for fault diagnosis, the ANN must first be trained using data that are representative of the normal condition and of the various fault conditions. The inputs are seven normalized values of the residuals in Equations 1 through 7 and the outputs are nine values that constitute a pattern that represents the normal mode or one of the eight fault modes of operation. Hence, nine input/output patterns are used to train the network. Actual measured data for normal operation may be available from historical databases or can be obtained as the system operates. However, this may not be the case for fault conditions. Introducing a fault to the system so that fault data can be collected may not be possible due to concerns for occupant comfort. Hence, an alternative method for obtaining patterns of residuals during fault modes of operation is needed.

In this study, idealized training patterns are specified by considering the dominant symptoms of each fault. Following the discussion in the "Fault Description" section, examples of dominant symptoms/residuals for several faults are:

IF Supply fan failure **THEN** Supply fan rpm is zero.
Supply air pressure is zero.
Supply fan control signal is maximum.
Flow difference between supply and return ducts is zero.

IF Return fan failure **THEN** Return fan rpm is zero.
Flow difference between supply and return ducts increases.
Return fan control signal is maximum.

IF Pump failure **THEN** Cooling coil valve control signal changes.
Cooling coil valve position changes.

IF Cooling coil control valve failure **THEN** Cooling coil valve control signal changes.
Cooling coil valve position does not change.

Using this type of reasoning it is possible to construct a pattern of dominant training residuals for each fault. The matching of dominant residuals to the various faults is depicted in Figure 5.

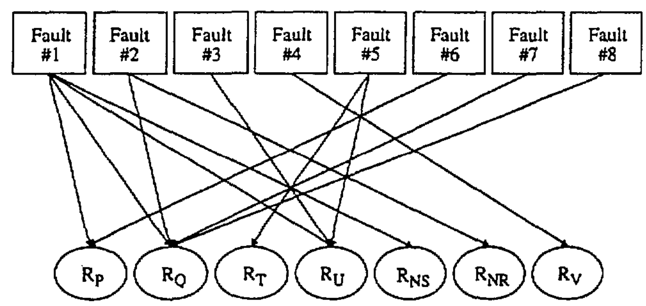

Figure 5 Matching of dominant residuals and faults.

The residuals are normalized so that the dominant symptom residuals have the same magnitude for the different fault cases. A dominant symptom residual is assigned a value of ±1 depending on the sign of the residual, and all other residuals are assigned a value of 0. The idealized input/output training patterns for the normal mode of operation and the eight faults are given in Table 1. The input patterns are based on conditions that are expected to exist after the system has reached steady state. Each output training pattern consists of eight values of 0 and one value of 1. The normal mode has a 1 as the output for the first unit, fault 1 has a 1 as the output for the second unit, and so on.

For testing of actual data, a normalized residual is obtained by dividing a residual from Equations 1 through 7 by the absolute value of the maximum value obtained for this residual from measured fault data. Thus, the maximum value of the absolute value of R_T obtained for a particular fault is used to normalize the supply air temperature residual for all the considered faults. The normalized residual for the supply air temperature is:

$$R_T = \frac{T_S - T_{S,SP}}{|T_S - T_{S,SP}|_{max}} \tag{13}$$

where the subscript "*max*" denotes maximum.

The ANN architecture is 7 × 5 × 9, where the first number is the number of inputs (residuals, n_i), the last number is the number of outputs (normal mode plus eight fault modes, n_o), and

TABLE 1 Normalized Patterns for AHU Fault Diagnosis Used in ANN Training

Net Inputs							Net Outputs									Fault Diagnosis
R_P	R_Q	R_T	R_U	R_{NS}	R_{NR}	R_V										
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	Normal
−1	−1	0	1	−1	0	0	0	1	0	0	0	0	0	0	0	#1 Supply fan
0	1	0	0	0	−1	0	0	0	1	0	0	0	0	0	0	#2 Return fan
0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	#3 Pump
0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	#4 Cooling coil valve
0	0	−1	−1	0	0	0	0	0	0	0	0	1	0	0	0	#5 Thermocouple
−1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	#6 Pressure transducer
0	−1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	#7 Supply flow station
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	#8 Return flow station

the middle number is the number of units in the hidden layer (n_h). A log-sigmoid activation function is used for both the hidden and output layers. The network is trained until the sum-of-squares error is less than 10^{-6} or until the number of training epochs exceeds 5,000. A commercial ANN software package was used for the training (Demuth and Beale 1992).

The methodology described above is one example of how ANNs can be used for fault detection and diagnosis. Model-based approaches are a second example of the use of ANNs for this task. Model-based approaches compare ANN models of normal and faulty system or subsystem operation to the actual system operation. Diagnosis of the current state of the system is based on determining which model has the greatest degree of similarity to the actual system operation. The model-based approach is not used in this study.

Network Topology and Training

The selection of the appropriate number of hidden layers and the number of units in a layer is problem dependent and typically requires considerable engineering judgment (Schalkoff 1992). As is the case for most numerical algorithms, a tradeoff between accuracy and computational requirements may exist. For instance, by adding more hidden units and layers to a network, the agreement between the actual and target outputs may be improved but at the cost of increased training time and memory requirements. In addition, if too many hidden units are used, overfitting of the training data may occur and the generalization to new input patterns may be poor. This is similar to the effect seen when curve-fitting with too many free parameters (Hertz et al. 1991).

Hecht-Nielson (1990) provides guidelines for training and testing ANNs. The basic approach is to divide the available data into a training set and a testing set. Both sets should include data that cover the full range of operating conditions. The amount of training required to yield a sufficiently accurate ANN is also problem dependent. For most ANNs an optimum number of training epochs exists that minimizes the error for the testing data. Additional training epochs will most likely yield lower training errors; however, the errors for the test data may increase. This phenomenon occurs commonly for ANNs trained with the backpropagation learning rule and is known as *overtraining*. Overtraining can result in an ANN that exhibits poor generalization because the ANN simply memorizes the input training patterns. In most cases, the appropriate network topology and number of training epochs can only be determined through an extensive trial-and-error process.

RESULTS AND DISCUSSION

Faults in the AHU are diagnosed by inputting residual vectors to the trained ANN. The residual vectors are obtained by introducing faults in the laboratory AHU and recording the subsequent response of the system. The system response is input in a batch mode to the trained ANN model of the cooling coil value subsystem to compute $U_{CC,EV}$. This computation could also be done on-line. Residuals are calculated using steady-state values of the system variables measured 900 seconds after a fault is introduced. Residuals for the normal mode and eight fault modes of operation are given in Table 2. Normalized residuals computed using expressions such as Equation 13 are given in Table 3 and are used as input to the ANN for fault diagnosis.

A more systematic approach to determining the steady-state residual values would be to develop and implement a steady-state detector. One possible method for detecting steady-state conditions would be to use regression techniques to obtain linear equations that characterize the responses of variables such as T_S, P_S, and Q_D. Steady-state conditions would be indicated if the slopes of these lines were less than the threshold values defined for each signal. A steady-state detector such as this would be necessary for on-line implementation in a real system because the onset of a fault is not known *a priori*.

Results of the fault diagnosis are given in Table 4. From the training patterns in Table 1, a perfect diagnosis would yield values on the diagonal of unity, and all other values would be zero. The values on the diagonal in Table 4 are near unity (underlined), indicating that the ANN successfully diagnosed each condition. Thus, although the training was based on simple, idealized relationships between the symptoms and faults, the ANN accurately discriminates between the various faults and the

TABLE 2 Measured Residuals 900 s After the Occurrence of a Fault

Fault	System Operation	R_P	R_Q	R_T	R_U	R_{NS}	R_{NR}	R_V
	Normal	−0.004	0.011	−0.080	0.130	0.037	0.065	0.011
#1	Supply fan fault	−0.249	−0.398	0.790	4.005	−9.983	0.017	−0.249
#2	Return fan fault	0.004	0.652	0.063	−0.097	0.071	−9.983	−0.042
#3	Pump fault	−0.004	0.079	0.013	2.145	0.043	0.072	−0.354
#4	Cooling coil valve fault	−0.003	−0.068	−1.095	−1.577	0.044	0.016	3.247
#5	Thermocouple fault	0.002	−0.027	−14.500	−3.697	0.063	0.056	−0.274
#6	Pressure transducer fault	−0.249	0.036	0.010	0.529	−0.003	0.079	−0.010
#7	Supply flow station fault	0.005	−0.868	−0.072	−0.086	0.055	0.017	0.014
#8	Return flow station fault	−0.001	1.072	0.000	−0.299	0.048	−0.011	0.099

TABLE 3 Normalized ANN Input for Fault Diagnosis

Fault	System Operation	R_P	R_Q	R_T	R_U	R_{NS}	R_{NR}	R_V
	Normal	−0.016	0.010	−0.006	0.032	0.004	0.007	0.003
#1	Supply fan fault	−1.000	−0.371	0.054	1.000	−1.000	0.002	−0.077
#2	Return fan fault	0.016	0.608	0.004	−0.024	0.007	−1.000	−0.013
#3	Pump fault	−0.016	0.074	0.001	0.536	0.004	0.007	−0.109
#4	Cooling coil valve fault	−0.012	−0.063	−0.076	−0.394	0.004	0.002	1.000
#5	Thermocouple fault	0.008	−0.025	−1.000	−0.923	0.006	0.006	−0.084
#6	Pressure transducer fault	−1.000	0.034	0.001	0.132	0.000	0.008	−0.003
#7	Supply flow station fault	0.020	−0.810	−0.005	−0.021	0.006	0.002	0.004
#8	Return flow station fault	−0.004	1.000	0.000	−0.075	0.005	−0.001	0.030

TABLE 4 Diagnosis Results for the Data in Table 3

System Operation	Output Pattern								
Normal	<u>1.000</u>	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Supply fan fault	0.000	<u>1.000</u>	0.000	0.000	0.000	0.000	0.022	0.000	0.000
Return fan fault	0.000	0.000	<u>1.000</u>	0.000	0.000	0.000	0.000	0.000	0.000
Pump fault	0.017	0.001	0.000	<u>0.927</u>	0.000	0.000	0.000	0.000	0.000
Cooling coil valve fault	0.004	0.000	0.000	0.000	<u>0.998</u>	0.000	0.041	0.001	0.000
Thermocouple fault	0.000	0.000	0.102	0.000	0.000	<u>1.000</u>	0.000	0.000	0.000
Pressure transducer fault	0.000	0.000	0.000	0.000	0.000	0.000	<u>1.000</u>	0.000	0.000
Supply flow station fault	0.005	0.000	0.000	0.000	0.000	0.002	0.000	<u>0.999</u>	0.000
Return flow station fault	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	<u>1.000</u>

TABLE 5 Additional Training Patterns for Select Faults

Net Inputs							Net Outputs									Fault Diagnosis
R_P	R_Q	R_T	R_U	R_{NS}	R_{NR}	R_V										
0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	#3Pump
0	0	0	0.5	0	0	0	0	0	0	1	0	0	0	0	0	#3Pump
0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	#4Cooling coil valve
0	0	0	0	0	0	0.5	0	0	0	0	1	0	0	0	0	#4Cooling coil valve

normal condition when actual data are used. Because of their ability to generalize and to filter noise, ANNs appear to be useful tools for fault diagnosis. For the set of faults and the associated symptoms considered in this study, fault diagnosis methods based on if-then rules, or pattern recognition techniques such as the nearest neighbor algorithm (Schalkoff 1992), are also expected to be effective. As the number of faults increases, however, implementation of the if-then rules may become cumbersome. Extension of the ANN method for fault diagnosis to include other faults is expected to be straightforward. The major computational requirement for ANNs occurs during training (which can be performed off-line) and this is not expected to present a problem during on-line operation.

It should be noted that the success achieved for this set of data is not guaranteed in general. Because training is based on a small set of idealized data, generalization problems can occur when actual data are considered. This potential difficulty is easily envisioned for normalized residuals with values near ±0.5, as for fault 3 in Table 3. Although this fault is correctly diagnosed, this set of residuals could also have been identified as a normal operating condition. In this case, the problem is linked to the severity of the fault. That is, a more severe pump fault would produce a larger value of U_{CC} and, therefore, a larger value of R_U. This would improve the likelihood of a correct diagnosis.

A second case where generalization may be imperfect relates to the state of the system when the fault occurs. As an example, consider the stuck-valve fault. If the valve sticks in a position where it is roughly half open and the value used for normalization is based on the maximum possible difference between the actual and expected positions, the corresponding normalized residual R_V will again be near ±0.5. Thus, the same kind of generalization problem as cited previously could be encountered.

The reliability of the ANN for diagnosing imperfect input data patterns can be improved in two ways. First, the input training data set can be extended to include patterns that account for less severe faults and faults related to the state of the system prior to the occurrence of the fault. For instance, the training patterns for the pump fault and valve fault could be extended as shown in Table 5. This kind of extension of the input training data is rather straightforward to envision once the patterns for the severe faults have been established. A second way in which the ANN can be taught to generalize more accurately is by training the network with noise added to the idealized input patterns. For instance, random noise that is normally distributed with a mean value of 0 and a variance of 0.1 can be added to the input patterns in Table 1. The training data then consist of the original idealized input patterns and the noisy patterns. Additional noisy input patterns with different values of the variance can also be added to the training data set. Both approaches are currently being investigated to improve the robustness of the diagnosis.

It is also possible that the input to the ANN will represent a fault mode of operation for which the ANN was not trained. In fact, it seems probable that this will occur occasionally and therefore must be accounted for in an actual implementation of the method. The most desirable output in this scenario would be a *warning that the system is operating in some unknown fault* mode. However, because the training data do not include this type of input, the output may be erroneous. Although this may appear to be a drawback of the ANN method, the reality is that this scenario is likely to cause problems regardless of the method used for fault diagnosis.

The diagnosis of experimental faults is based on steady-state or near-steady-state conditions and dynamic conditions are not considered. Further study is needed to determine how the method can be extended for use when dynamic conditions exist.

CONCLUSIONS AND RECOMMENDATIONS

The objective of this study was to describe the application of ANNs to the problem of fault diagnosis in an AHU. Initially, residuals of system variables were selected that could be used to quantify the dominant symptoms of fault modes of operation. Idealized steady-state patterns of these residuals were then defined for each mode of operation studied. The steady-state relationship between the dominant symptoms and the faults was learned by an ANN using the backpropagation algorithm. The trained ANN was applied to experimental data for various faults and successfully identified each fault.

An ANN was also used successfully as a model for a cooling coil valve subsystem. The output of the ANN was the expected value of the cooling coil valve control signal. Although the agreement between the actual and predicted control signals during normal operation was not perfect, the ANN model was adequate for identifying normal and fault modes of operation. The agreement for normal operating conditions could be improved by changing the inputs to the ANN; however, this may lead to a model that tracks the operation of the valve during fault conditions rather than providing an estimate of the normal operation of the valve under normal operating conditions.

This study demonstrates the feasibility of using ANNs for diagnosis of faults in HVAC systems. Eight fault modes of operation were considered and all faults were of a severe nature. Hence, the symptoms of these faults are relatively easy to distinguish. Nonetheless, it is anticipated that, because of their abilities to learn complex, nonlinear relationships and to generalize, ANNs will also be effective for less severe faults.

The method can be extended in a straightforward manner to consider additional faults such as damper faults in the mixing box. It is likely that this will require the introduction of additional residuals to the analysis. As the complexity of the system and the number of faults considered grows, it may be desirable to develop separate ANNs for various subsystems and to use a preprocessor to identify the appropriate subsystem to examine for the existence of a fault.

ACKNOWLEDGMENTS

This research was partially supported by the Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy.

REFERENCES

Ananthraman, S. 1995. Intelligent process optimization on the plant floor. *Industrial Computing* 14(4): 40-44.

Curtiss, P.S., J.F. Kreider, and M.J. Brandemuehl. 1993. Energy management in central HVAC plants using neural networks. *ASHRAE Transactions* 99(1): 496-504.

Demuth, H., and M. Beale. 1992. *Neural network toolbox: User's guide*. Natick, Mass.: The MathWorks, Inc.

Fan, J.Y., M. Nikolaou, and R.E. White. 1993. An approach to fault diagnosis of chemical processes via neural networks. *AIChE Journal* 39(1): 82-88.

Hecht-Nielson, R. 1990. *Neurocomputing*. Reading, Mass.: Addison-Wesley.

Hertz, J., A. Krogh, and R.G. Palmer. 1991. *Introduction to the theory of neural computation*. Reading, Mass.: Addison-Wesley.

Koive, H.N. 1994. Artificial neural networks in fault diagnosis and control. *Control Eng. Practice* 2(1): 89-101.

Lee, W.Y., C. Park, and G.E. Kelly. 1996. Fault detection of an air-handling unit using residual and recursive parameter identification methods. *ASHRAE Transactions* 102(1).

Schalkoff, R.J. 1992. *Pattern recognition: Statistical, structural, and neural approaches*. New York: John Wiley & Sons.

Watanabe, K., S. Hirota, L. Hou, and D.M. Himmelblau. 1994. Diagnosis of multiple simultaneous faults via hierarchical artificial neural networks. *AIChE Journal* 40(5): 839-848.

QUESTIONS AND COMMENTS

Jean Lebrun, Professor, Laboratoire de Thermodynamique, University of Liege, Liege, Belgium: I guess the ANN technique could also be used to interpret time variations: we could include time derivatives as input and output variables.

John House: You bring up a very good point. Yes, time derivatives could be included as inputs to the neural network, although we have not attempted this. Including time derivatives may make it possible to diagnose faults more quickly by eliminating the need to wait for steady-state conditions to prevail. In certain circumstances this may be advantageous, but in most cases it would seem to introduce additional questions. For instance, can we expect the dynamic response of a given system to always be the same for a given fault or is it dependent on the operating point or load? Can we expect the response to be the same in another system? Are we better off to wait for steady-state conditions to exist since the faults are not typically life threatening and we have a better sense of the state to which the system will evolve for a given fault?