

Causality-Based Failure-Driven Learning in Diagnostic Expert Systems

It has been recognized that a diagnostic expert system's ability to learn from past experience will improve its diagnostic efficiency as well as make it acquire new heuristics. In this paper, we propose a failure-driven learning scheme by which the expert system automatically updates its compiled knowledge by acquiring new heuristics or refining existing heuristics. A heuristic is refined if it hypothesizes the wrong causal origin during a diagnosis. Using its deep-level knowledge of the process, the expert system draws inductive inferences from causal models to determine why the hypothesis proposed by the heuristic is inconsistent with the current state of the process. The refinement limits the applicability of the heuristic and prevents it from firing if a similar situation were to subsequently arise.

Steven H. Rich
Venkat Venkatasubramanian
Intelligent Systems in Process Engineering
Laboratory
Department of Chemical Engineering
Columbia University
New York, NY 10027

Introduction

A fault in a chemical process may be defined as "a departure from an acceptable range of an observed variable or calculated parameter associated with the equipment" (Himmelblau, 1978). Fault diagnosis can be defined as a search for the causal origin of an observed pattern of abnormal system behavior. For chemical processes, the causal origin might be a process unit malfunction, an operator error, or variability in some raw material used by the process. Automating fault diagnosis by means of encoding diagnostic knowledge in an expert system has recently gained widespread interest. A diagnostic expert system can be evaluated based on the speed with which it reaches its conclusions and the confidence with which these conclusions can be regarded by the system user. To enhance diagnostic speed, heuristics, representing compiled knowledge from past experience, are included in a knowledge base to propose direct associations between system observations and system device states. The diagnostic reliability of an expert system which draws conclusions solely from heuristics suffers in that the system has no means of verifying the intermediate causal relationships skipped by the heuristics. Moreover, these diagnostic systems tend to be brittle, with the expert system failing abruptly when asked to locate system malfunctions for which no available heuristic applies. To improve diagnostic reliability, device models, which consider device structure and functionality, can be added to the knowl-

edge base (Rich and Venkatasubramanian, 1987; Venkatasubramanian and Rich, 1988). By relying upon device models and device interconnections, a diagnostic expert system can locate a disturbance origin (e.g., a device malfunction) by searching for a causal path along which the disturbance propagates its effects. In this manner a link is established between a system disturbance and the observed system abnormalities. A diagnosis relying solely on deep-level knowledge tends to be slower relative to a diagnosis based on compiled knowledge because in its causal path search the expert system proceeds methodically from one device to a neighboring device, attempting to chain together a set of local causal relationships.

In this trade-off between speed and reliability, a strategy must be devised to construct a tier of compiled knowledge that will strike a balance among completeness, heuristic generality, and heuristic specificity. A strategy, based on machine learning concepts, in which the compiled knowledge base is developed incrementally by generating new heuristics and refining old ones as the system accumulates diagnostic experience appears to be one alternative. New heuristics are generated when no heuristic is found in the current knowledge base that is applicable to the current event (or set of events) being diagnosed. Heuristics are refined either by rule specialization or rule generalization. A heuristic is specialized by including additional premises to the existing heuristic in order to limit its applicability. For rule generalization, heuristics are combined to reduce redundancy in the knowledge base. In this paper we propose a learning scheme for acquisition of new heuristics and specialization of existing ones in which the refinement procedures are based on causal knowl-

Correspondence concerning this paper should be addressed to V. Venkatasubramanian, Laboratory for Intelligent Process Systems, Department of Chemical Engineering, Purdue University, W. Lafayette, IN 47907.

edge found in the deep-level tier of the two-tier knowledge base.

The role of learning is easily appreciated in the case of future generation knowledge-based systems for diagnosing faults in large complex processes, where the expert system's knowledge base for some new process will be created by assembling causal and fault models of the process units from a library that was created in the past. Then by driving this deep-level model-based diagnostic expert system by a process simulator that simulates the process behavior under different fault conditions, the system can recognize important patterns and compile them as heuristics. For such an acquisition of compiled knowledge, learning strategies such as the failure-driven learning technique proposed in this paper would prove to be useful.

This paper is divided into two principal sections. The first section examines several inductive learning strategies for acquiring and refining heuristics. The proposed learning strategy based on causal models is presented in the second section. Three sample problems are offered to illustrate the approach.

Fault Diagnosis and Failure-Driven Learning

Inductive learning (Michalski, 1983) is a process of acquiring knowledge in which hypotheses are drawn from observations that represent some knowledge about an object, situation, or process. The induced hypotheses, which attempt to provide an explanation or a description of these observations, must be consistent with the observations and satisfy any assumptions or constraints known to be imposed on the observations. Learning from examples and learning from experience represent two special cases of inductive learning. Mitchell (1982) and Mitchell et al. (1983) have proposed an inductive learning strategy to develop a rule base in a learning program that automatically generates its practice problems. An intelligent strategy for generating these problems must be incorporated within the program to ensure that the practice problems are illustrative of a wide variety of concepts. To avoid this complication, Politakis and Weiss (1984) suggest using a set of case histories with known expert interpretations. They have adapted the version state learning approach to revising a set of diagnostic production rules derived from empirical observations.

Unlike diagnoses in domains in which the bulk of the relationships between symptoms and causes can only be established on an empirical basis, the association between atypical process behavior and process malfunctions can be inferred from device models. Recognizing this fact, researchers who have proposed learning strategies for refining and acquiring heuristics for process diagnosis have suggested using deep-level knowledge at the core of the system. Before continuing with a description of these learning systems, we will first turn our attention to a formal definition of diagnosis as developed by Reiter (1987). Reiter begins with

1. A system comprised of a set of m structures $\{s1, \dots, sm\}$ and a set of n components $\{c1, \dots, cn\}$; structures represent subsets of components at a higher level of aggregation
2. A system description SD
3. A set of observations OBS based on system behavior and introduces the notation $AB(.)$ to represent a predicate interpreted to mean "abnormal." The set $\{\sim AB(c1), \dots, \sim AB(cn)\}$ represents the assumption that all system components are behaving correctly. Reiter then concludes that a malfunction is

present in the system if the implication

$$SD \text{ and consistent } [\{\sim AB(c1), \dots, \sim AB(cn)\}] \rightarrow OBS$$

cannot be made where consistent (H) is true if believing the hypothesis H does not result in a contradiction. In other words, the implication cannot be made because the set of system observations OBS conflicts with the expected typical behavior of the system. Reiter notes that consistency can be restored if a subset of the assumptions $\sim AB(s1), \dots, \sim AB(sm), \sim AB(c1), \dots, \sim AB(cn)$ are retracted (retracting an assumption associated with a structure provides diagnostic focus) and defines a diagnosis for $(SD, COMPONENTS, OBS)$ as the minimal set Δ of system components such that the implication

$$SD \text{ and consistent } [\{AB(c) | c \in \Delta\} \text{ and } \{\sim AB(c) | c \in COMPONENTS - \Delta\}] \rightarrow OBS$$

can be made.

In performing a diagnosis using compiled knowledge, we are looking for a set of heuristic rules to fire that will establish the set Δ . If the set of components Δ conjectured by the heuristics does not restore consistency then the heuristics are said to have failed either because they proposed that a component was in an abnormal state when in fact it was behaving correctly or the set of heuristics was incomplete and consequently the entire set Δ could not be enumerated. This represents a negative training instance and so the heuristic rule base must be revised both by acquiring new heuristics to fill gaps in the knowledge base, and by strengthening heuristics which offered wrong conclusions so that the heuristic would not be applicable given a similar training instance. Learning in this manner has been termed *failure-driven learning*.

Failure-driven learning: one approach

Two such failure-driven learning approaches for process diagnosis have been proposed, the first by Steels and Van de Velde (1985) and the second by Pazzani (1986). Steels' heuristic rules (structured as IF/THEN production rules) have three parts:

1. Primary symptoms representing the rule premise prior to refinement
2. Secondary symptoms representing the conditions added to a rule during the refinement stage
3. Rule conclusion representing the basic event to which the primary symptom can be attributed.

If the current rule base fails to arrive at a correct diagnosis, under a particular set of circumstances, then the system falls back upon a core of first-principles knowledge encoded as a set of fault trees corresponding to the process. The following rule integration scheme is proposed to update the heuristic rule base in order, first, to revise the rule that proposed the wrong conclusion, and second, to add a new rule, which under similar circumstances would fire to propose the proper diagnosis. The scheme is:

1. Revise the old rule by adding secondary symptoms that represent the negation of the conclusion of the new rule
2. Add a new rule of the form:
 - A primary symptom corresponding to the top-level event that triggered the diagnosis
 - Secondary symptoms that are void
 - A rule conclusion corresponding to the basic event that

first-principles knowledge has concluded causes the primary symptom.

The construction of the new rule is such that all rules remain mutually exclusive, that is, no two rules should fire and propose a different rule conclusion for the same primary symptom. The second method for updating the heuristic rule base is rule combination, which is used to reduce the size of the rule base when two rules have a common conclusion. The authors argue that rule combination improves diagnostic efficiency in that system conditions need not be checked repeatedly. The argument appears to be weak because most expert systems have a means of caching information as it becomes available so that conditions need only be checked once. Moreover, repeated investigation of conditions can also be avoided by reordering a rule base so as to sequence the order in which rules can fire.

Example

A diagnosis of a simple process comprised of a heat exchanger connected to a storage tank, shown in Figure 1, is presented to illustrate Steels' learning strategy. Assume that the initial rule set consists of the following two heuristics:

Rule 1

IF tank level is low
THEN propose that pipe 1 is leaking

Rule 2

IF stream 2 temperature is low
THEN propose that pipe 1 is leaking

If an observation is made that the temperature of stream 2 is low and the diagnosis reveals that the cause is a low stream 3 temperature, then the rule set is revised via rule integration as follows:

Rule 1

IF tank level is low
THEN propose that pipe 1 is leaking

Revised Rule 2

IF stream 2 temperature is low
and stream 3 temperature is not low
THEN propose that pipe 1 is leaking

New Rule 3

IF stream 2 temperature is low
THEN propose that stream 3 temperature is low

Rule combination of rules 1 and 2 can now be used to further reorganize the rule base:

Rule 1

IF tank level is low
and stream 2 temperature is low
THEN propose that pipe 1 is leaking

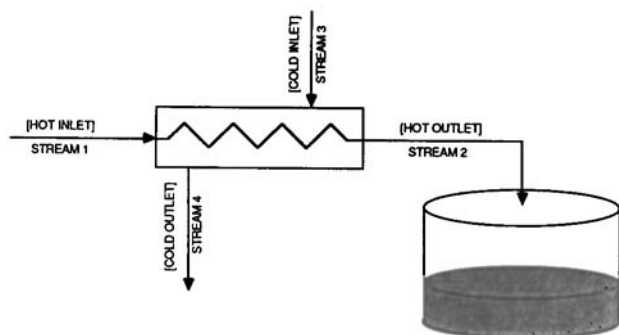


Figure 1. A simple process.

Rule 2

IF stream 2 temperature is low
THEN propose that stream 3 temperature is low

Failure-driven learning: a second approach

Pazzani has proposed a more elaborate failure-drive learning scheme in which first-principles knowledge is represented in the form of device models. For example, the simple heuristic

IF sensor is reading zero
THEN the sensor is faulty

can be revised to include conditions that distinguish a sensor malfunction from a disturbance to the control variable which the sensor monitors. The revision might take the form:

IF sensor reads zero
and the controller (to which the sensor sends its signal)
does not compensate for the zero reading
THEN the sensor is faulty

Pazzani had identified three reasons for hypothesis failure. These are summarized below.

Hypothesized Fault—Inconsistent Prediction. In this case the fault, hypothesized as the basic event to which a top-level event can be attributed, is inconsistent with observed system behavior. The correction strategy for refining the heuristic requires a check for system constraints that the fault might violate (or satisfy) if its effects were propagated through the system. Consider a simple process with two streams and a valve. Call the inlet stream, stream 1, and the outlet stream, stream 2. If the valve is manually controlled then the following heuristic might be applicable in diagnosing a low stream 2 flow rate

IF stream 2 flow rate is low
THEN propose that valve position setting is incorrect

The hypothesis would be rejected if the current state of the system shows that stream 1 flow rate is greater than stream 2 flow rate, that is, the material balance around the valve is violated. One possible explanation is a valve leak. The heuristic is revised accordingly:

IF stream 2 flow rate is low
and stream 1 flow rate = stream 2 flow rate
THEN propose that valve position setting is incorrect

If the material balance is satisfied, then the updated heuristic distinguishes between a valve leak and operator error.

Hypothesized Unusual Mode—Enablement Violated. For this case, (apparent) abnormalities are hypothesized to arise from the system being in a normal but unusual mode. However, the enabling conditions that must be satisfied for the system to be in this mode are not met. The correction strategy is to consider an enabling condition for the unusual state. For example, if the level in a reactor is low, a heuristic rule might propose that this is a result of the system being in PUMPOUT mode in which the valves controlling the flow of the reactor inlet streams are closed and the valve regulating the flow of the reactor outlet stream is wide open. One possible enablement condition for PUMPOUT mode is that the reactor is being readied to be cleaned. The revised heuristic reads.

IF reactor level is low
and reactor is to be cleaned
THEN propose that low level is a result of the system
being in PUMPOUT mode

This heuristic distinguishes between the unusual operating mode and a faulty device state, for example, a tank leak.

Hypothesized Fault—Unusual Input. Rules that fall under this category propose that a device is faulty. A hypothesis failure results from the system being confused by an unusual input to the device that causes an unusual output. The correction strategy is to consider the device's functionality. As an example, suppose that a reactor's level is controlled by a negative feedback loop and the level is regulated by the flow of an inlet stream to the reactor. If the reactor level is low, a rule might fire proposing that the level controller is broken. The level controller can be checked by noting whether the control valve has been positioned to increase the flow rate in the inlet stream. The revised heuristic takes the form:

IF reactor level is low or high
and valve position is not abnormal in the direction
opposite of the reactor level

THEN propose that the level controller is broken

This revised heuristic discriminates between a broken controller and an external disturbance to the manipulated variable of the control loop.

Pazzani's device models are not described in sufficient detail to have a thorough understanding of the automatic learning mechanisms used to update heuristics. Therefore, a rigorous comparison of Pazzani's procedure and the learning strategy proposed in the next section cannot be made. One important distinction, however, is possible. Pazzani chooses to refine a heuristic immediately after it fails in contrast to the proposed approach in which heuristics are flagged when they fail and refined only at the conclusion of the diagnosis. It can be argued that Pazzani's approach might ultimately slow the diagnosis especially if the information gained while searching for the cause of the heuristic's failure proves to be of little value later in the diagnosis.

Heuristic Acquisition and Refinement Based on Casuality

In this section a new failure-driven learning strategy is proposed for acquiring new diagnostic heuristics and strengthening existing heuristics in the compiled knowledge base of a two-tier architecture. The learning scheme is to be used as a means for filling gaps in the compiled knowledge base and reducing the likelihood of heuristics firing and hypothesizing a causal source that is incorrect. As noted earlier, by employing a learning scheme to update the compiled knowledge tier we are endeavoring to improve diagnostic efficiency such that the bulk of any diagnosis is performed using heuristic knowledge with the system dropping down to its first-principles knowledge either:

1. To diagnose events (or patterns of events) that occur infrequently, or

2. To reach a diagnostic conclusion at a causal source (e.g., a device malfunction) whose probability of occurrence is low

Two other factors were also taken into account in devising the learning strategy. First, it was recognized that a large number of case studies may not be available to be used as training examples for process diagnosis (e.g., on a plant retrofit or plant expansion). Hence, the compiled knowledge base should be progressively refined with each diagnostic consultation performed by the expert system. Second, the learning strategy should fit comfortably in the existing MODEX2 architecture (Venkatasubramanian and Rich, 1988), making use of the available library of

fault and causal models to reach decisions on how best to revise heuristics and acquire new ones.

In the diagnosis of a process abnormality in the current version of MODEX2, control initially passes to the compiled knowledge base. A diagnostic method (attached to the object in which the abnormality has occurred) identifies certain heuristics as potentially being applicable. If a heuristic's premises are satisfied, then the heuristic suggests to the system user to confirm or deny the presence of the basic event stored in the heuristic's conclusion. If the presence of the basic event is confirmed, then the expert system presents a diagnostic summary. At this point the system user is offered the opportunity to drop down to first-principles knowledge in order:

1. To verify all intermediate causal relations omitted in the heuristic, and

2. To search for additional potential causes of the abnormality

If the heuristic fails, that is, it proposes a hypothesis that proves to be false, then MODEX2 looks to the next heuristic. If no heuristic is found to be currently applicable the system drops down to first principles and considers a fault model in order to locate the local causal origin of the process abnormality. If this local causal origin is not a basic event, then diagnostic attention is focused on this new fault and control passes back to the compiled knowledge base as directed by the appropriate method. The diagnostic process continues in this recursive manner until either a heuristic succeeds in locating the root cause of the initially observed process abnormality, the root cause is found using the first-principles knowledge, or all potential diagnostic lines of reasoning as offered by the first-principles knowledge base are exhausted.

With the learning scheme in place, a diagnosis proceeds as described above with the exception that the expert system keeps track of all heuristics that fired and proposed incorrect hypotheses during the course of the diagnosis. The system makes note of which basic events were incorrectly hypothesized and, using causal models, the system determines how the effects of these events propagate locally in the process.

At the end of the diagnosis, the system checks whether these local causal effects established by the causal model are contradicted. For example, if a heuristic fires and incorrectly proposes that a valve is clogged then the expert system checks whether the flow rates of the inlet and outlet streams from the valve have been considered in the course of the diagnosis. If this is the case and one of the flow rates is found not to be low then a contradiction is established. Based on the presence or absence of a contradiction at the conclusion of the diagnosis, the expert system proposes the following updating scheme for the compiled knowledge base:

1. If a contradiction is found, then strengthen the heuristic to which the contradiction is associated, by including the local causal effect as an additional premise.

2. Add a new heuristic to the compiled knowledge base of the following form:

Case 1: No contradiction found

IF process abnormality that triggered diagnosis is present and during course of shallow-level diagnosis negates conclusion(s) of any heuristics that fired

THEN propose basic event to which process abnormality can be attributed in this particular example

Case II: Contradiction found

IF process abnormality that triggered diagnosis is present
THEN propose basic event to which process abnormality can be attributed in this particular example

Note that the rule inclusion method proposed in this section differs from the rule integration approach of Steels and Van de Velde in that secondary symptoms, representing the negation(s) of rule conclusion(s), are added to the new heuristic, not the heuristic that fired during the diagnosis. This strategy is based on the assumption that the original set of heuristics in the knowledge base associates process abnormalities with those basic events that will occur with the greatest frequency. In practice, these are the first heuristics likely to be proposed by a domain expert. Consequently, it would not appear advantageous to strengthen these original heuristics by appending negations of other heuristic conclusions as Steels and Van de Velde propose, for this would imply that the system user would examine basic events in the reverse order of their likelihood of occurrence.

This rule inclusion procedure represents an effort to achieve a compiled knowledge base with a set of mutually exclusive rules. In case I, in which no contradiction is found, the "negate heuristic conclusion" premise is used to impose an ordering over which diagnostic conclusions are proposed. If a contradiction is found, the update of the existing heuristic with a new premise describing the local causal effect discriminates between the existing heuristic and the new heuristic (acquired using case II above). When heuristics have been compiled from simulations, contradictions will be unlikely as the simulation will have considered the propagation of the local causal effect of the basic event. However, learning can identify sets of heuristics that are not mutually exclusive and the rule inclusion update strategy shown in case I above will prove useful in establishing an ordering over which these heuristics are searched.

It is likely that a failure-driven learning strategy will need to be used in tandem with a learning mechanism that is invoked when the diagnosis is successful. When diagnoses are successful, statistics can be compiled in an effort to determine the underlying probability distribution of event occurrences. If these statistics are considered in conjunction with a review of the knowledge base updates resulting from the failure-driven learning, the knowledge base can be slightly reorganized to resequence the order in which heuristics are searched. For example, heuristic premises of the type "negate heuristic conclusion" (as used in the rule inclusion strategy when no contradiction is found) can be repositioned among the heuristics. This repositioning should affect a new ordering of heuristic firings that is consistent with the relative frequencies of event occurrences.

Case studies

Three sample examples of rule base revisions as based on this proposed learning scheme will now be presented. In the first two examples, the resulting rule base will be compared to an updated rule base based on the approach of Steels and Van de Velde. The first sample problem to be considered is that of the heat exchanger and storage tank shown in Figure 1. The initial rule set reads

Rule 1

IF tank level is low
THEN propose that pipe 1 is leaking

Rule 2

IF stream 2 temperature is low
THEN propose that pipe 1 is leaking

If the temperature of stream 2 is found to be low then rule 2 fires, suggesting that the system user check whether pipe 1 is leaking. We will assume here that pipe 1 is free of leaks and since no other heuristics are applicable, MODEX2 would drop down to its deep-level knowledge and, based on the method for diagnosing low stream temperatures, the expert system would examine the temperatures and flow rates of stream 1 and stream 3. In finding that the flow rate of stream 1 is normal a contradiction is reached with the local causal effect of pipe 1 leakage, namely, that a leakage would result in a low stream 1 flow rate (MODEX2 finds this information in the causal model associated with pipes). The cause of the low stream 2 temperature is determined to be a low stream 3 temperature, which is considered to be a basic event in this example since the source of the material flowing in stream 3 is not within the boundaries of the system under study. The rule base is updated as follows with a revision of the second rule and the addition of a new, third rule:

Rule 1

IF tank level is low
THEN propose that pipe 1 is leaking

Revised Rule 2

IF stream 2 temperature is low
and stream 1 flow rate is low
THEN propose that pipe 1 is leaking

New Rule 3

IF stream 2 temperature is low
THEN propose that stream 3 temperature is low

In comparison to Steels' strengthening of the second rule with the statement "stream 3 temperature is not low," it appears that a more judicious selection of an additional rule premise has been made here since there is no causal relationship between the temperature of stream 3 and a pipe leak.

The second example to be considered is that of a reactor with one feed stream and one outlet stream, whose respective flow rates are regulated by manually controlled valves. The flowsheet is given in Figure 2. The revision of the tier of compiled knowledge is followed from the initial rule set of a single heuristic through three successive updates based on three separate diagnoses. The righthand columns in Table 1-3 show the revisions as guided by the causality-based learning scheme proposed in this paper. The lefthand columns list the revisions as guided by the approach presented by Steels and Van de Velde (1985).

Several interesting comparisons can be made. Note that in the third diagnosis as based on the learning scheme of Steels and

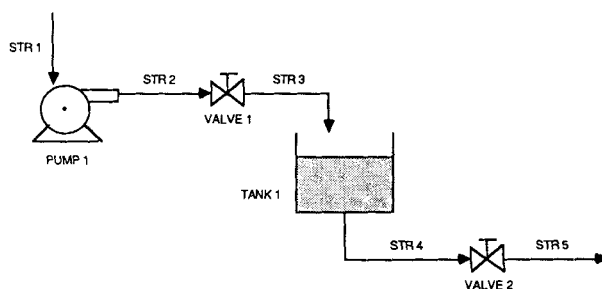


Figure 2. Flowsheet of case study 2.

Table 1. Analysis for Case Study 2—Part I

Diagnosis #1, ref. Figure 2 Process Abnormality: High tank level Cause: Wrong valve 2 position	
Approach of Steels & Van de Velde (1985)	This Paper Causality-Based Approach
Initial Rule Set	
Rule 1 IF tank level is high THEN propose that pipe 4 is blocked	Rule 1 IF tank level is high THEN propose that pipe 4 is blocked
Diagnosis	
1. Rule 1 fires and proposes that pipe 4 is blocked 2. Pipe 4 is blocked; drop down to first-principles diagnosis 3. High tank level attributed to wrong valve 2 position 4. Revise heuristic rule base as follows:	1. Rule 1 fires and proposes that pipe 4 is blocked. 2. Pipe 4 is not blocked; drop down to first-principles diagnosis • While diagnosing using deep-level knowledge check for a contradiction with low stream 4 flow rate (local causal effect of pipe 4 blockage) 3. High tank level attributed to wrong valve 2 position • No contradiction found 4. Revise heuristic rule base as follows:
Revised Rule 1 IF tank level is high and value 2 is <i>not</i> closed THEN propose that pipe 4 is blocked	Rule 1 IF tank level is high THEN propose that pipe 4 is blocked
New Rule 2 IF tank level is high THEN propose that valve 2 is closed	New Rule 2 IF tank level is high and pipe 4 is <i>not</i> blocked THEN propose that valve 2 is closed

Van de Velde the system user is asked to check for the presence of three basic events, wrong valve 1 position, wrong valve 2 position, and pipe 4 blockage, prior to the expert system's dropping down to first principles. In contrast, the revised knowledge base using the causality-based learning scheme asks the system user to check the value of one state variable (stream 4 flow rate) and to check for the presence of one basic event (wrong valve 1 position). Hence, after the second rule base revision, the expert system is able to diagnose with greater efficiency using the revised compiled knowledge base constructed with the causality-based learning scheme than with the rule base updated using the approach of Steels and Van de Velde. Note also the manner in which rules are strengthened in the left column. The original rule, which related a high tank level to pipe 4 blockage, now has three additional premises. If this heuristic had been proposed initially because pipe 4 blockage was the most likely event that could lead to a high tank level, then the addition of these new rule premises would have been counterintuitive; that is, the revised rule suggests that in diagnosing a high tank level three other potential causes should be checked prior to checking on the cause with the greatest likelihood.

Table 2. Analysis for Case Study 2—Part II

Diagnosis #2, ref. Figure 2 Process Abnormality: High tank level Cause: Wrong valve 1 position	
Approach of Steels & Van de Velde (1985)	This Paper Causality-Based Approach
Initial Rule Set: revised rule set from diagnosis #1, Table 1	
Diagnosis	
1. Check value 2 position and find that valve is not closed 2. Rule 1 fires and proposes that pipe 4 is blocked 3. Pipe 4 is not blocked; drop down to first principles 4. High tank level attributed to open valve 1 5. Revised heuristic rule base as follows:	1. Rule 1 fires and proposes that pipe 4 is blocked 2. Pipe 4 is not blocked; rule 2 fires 3. Valve 2 is not closed; drop down to first principles • While diagnosing look for contradiction with: (a) lock casual effect of pipe 4 blockage: low stream 4 flow rate (b) local casual effect of closed valve 2: low stream 4 flow rate, low stream 5 flow rate 4. High tank level attributed to open valve 1 • Contradiction found: stream 4 flow rate is high 5. Revise heuristic rule base as follows:
Revised Rule 1 IF tank level is high and valve 2 is <i>not</i> closed and value 1 is <i>not</i> open THEN propose that pipe 4 is blocked	Revised Rule 1 IF tank level is high and stream flow rate is low THEN propose that pipe 4 is blocked
Revised Rule 2 IF tank level is high and valve 1 is <i>not</i> open THEN propose that valve 2 is closed	Revised Rule 2 IF tank level is high and pipe 4 is <i>not</i> blocked and stream 4 flow rate is low THEN propose that valve 2 is closed
New Rule 3 IF tank level is high THEN propose that valve 1 is open	New Rule 3 IF tank level is high THEN propose that valve 1 is open

The final example, shown in Figure 3, is a modified version of the flowsheet of Figure 2 in which the level of the reactor is regulated by a negative feedback loop. Suppose that the initial version of the compiled knowledge base consists of the following single heuristic:

IF reactor level is high
 THEN propose that level set point is high (operator error)

If in diagnosing a high reactor level the heuristic fires and fails because the set point is at its proper value, then the expert system drops down to its first-principles knowledge. At the end of the diagnosis, checks are made for contradictions with the local causal effects of a high set point, namely, valve 2 closed and

Table 3. Analysis for Case Study 2—Part III

Diagnosis #3, ref. Figure 2 Process Abnormality: High tank level Cause: Improper pump speed	
Approach of Steels & Van de Velde (1985)	This Paper Causality-Based Approach
Initial Rule Set: revised rule set from Diagnosis #2, Table 2	
<ol style="list-style-type: none"> 1. Check valve 2 position and find that valve is not closed 2. Check valve 1 position and find that valve is not open 3. Rule 1 fires and proposes that pipe 4 is blocked 4. Pipe 4 is not blocked; drop down to first principles 5. High tank level attributed to high pump speed 6. Revise heuristic rule base as follows: 	<ol style="list-style-type: none"> 1. Check stream 4 flow rate and find that it is not low 2. Rule 3 fires proposing that valve 1 is open 3. Valve 1 is not open; drop down to first principles <ul style="list-style-type: none"> • While diagnosing look for contradiction with local causal effects of open valve 1: high stream 2 and stream 3 flow rates 4. High tank level attributed to high pump speed; no contradiction found 5. Revise heuristic rule base as follows:
<p>Revised Rule 1 IF tank level is high and valve 2 is <i>not</i> closed and valve 1 is <i>not</i> open and pump speed is <i>not</i> high THEN propose that pipe 4 is blocked</p>	<p>Revised Rule 1 IF tank level is high and stream 4 flow rate is low THEN propose that pipe 4 is blocked</p>
<p>Revised Rule 2 IF tank level is high and valve 1 is <i>not</i> open and pump speed is <i>not</i> high THEN propose that valve 2 is closed</p>	<p>Revised Rule 2 IF tank level is high and pipe 4 is <i>not</i> blocked and stream 4 flow rate is low THEN propose that valve 2 is closed</p>
<p>Revised Rule 3 IF tank level is high and pump speed is <i>not</i> high THEN propose that valve 1 is open</p>	<p>Revised Rule 3 IF tank level is high THEN propose that valve 1 is open</p>
<p>New Rule 4 IF tank level is high THEN propose that pump speed is high</p>	<p>New Rule 4 IF tank level is high and valve 1 is <i>not</i> open THEN propose that pump speed is high</p>

stream 4 flow rate responds to level disturbances (although set point is high, control loop is still assumed to be functioning properly). If the high tank level is a result of manual valve 1 being left wide open and the consequent saturation of the level control loop action, then a contradiction is found (valve 2 will be wide open) and the rule base is updated as follows:

Revised Rule 1

IF reactor level is high and valve 2 position is reduced past its nominal position corresponding to the correct set point and stream 4 flow rate varies in response to level disturbances

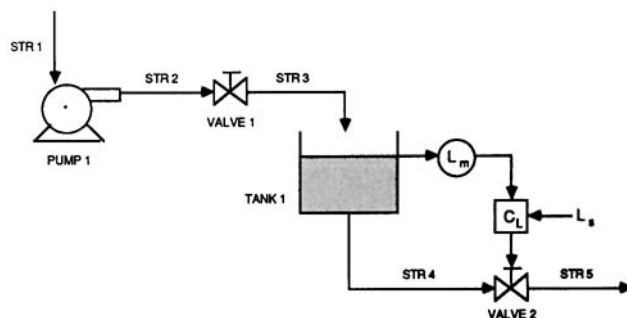


Figure 3. Flowsheet of case study 3.

THEN propose that level set point is high
New Rule 2

IF reactor level is high

THEN propose that valve 1 is wide open and level control loop action is saturated

If, on a second occasion, a high tank level is found to be the result of valve 2 being stuck closed, thereby inactivating the level control loop, then the rule base would be updated based on the contradiction with the local causal effect of valve 1 wide open. Note that rule 1 would not fire since the third rule premise of rule 1 is not satisfied; the flow rate of stream 4 would not change in response to reactor level disturbances because the control loop has zero gain.

Rule 1

IF reactor level is high and valve 2 position is reduced past its nominal position corresponding to the correct set point and stream 4 flow rate varies in response to level disturbances to the reactor level

THEN propose that level set point is high

Revised Rule 2

IF reactor level is high and stream 3 flow rate is high

THEN propose that valve 1 is wide open and level control loop action is saturated

New Rule 3

IF reactor level is high
THEN propose that valve 2 is stuck closed

Conclusions

In future generation diagnostic expert systems, the ability to learn from past experiences and compile away heuristic knowledge will prove to be a necessary and useful characteristic. To develop user satisfaction and confidence, a diagnostic expert system must diagnose quickly and reliably. These two system attributes can potentially be in conflict such that system modifications introduced to improve system performance relative to diagnostic speed may have an adverse effect on diagnostic reliability. We then look toward approaches to balance this trade-off between speed and reliability. With a two-tier knowledge base structure in which a tier of compiled knowledge is supported by a lower tier of deep-level knowledge, one endeavors to improve diagnostic efficiency such that the bulk of any diagnosis is performed reliably using compiled knowledge with the system dropping down to its deep-level knowledge either to diagnose events (or patterns of events) that occur infrequently, or to reach

a diagnostic conclusion at a causal source (e.g., a device malfunction) whose probability of occurrence is low. Such a two-tier framework is also a better mental model of a human operator's diagnostic reasoning process.

The role of learning is also easily recognized in the case of knowledge-based systems for diagnosing faults in large complex processes, where the expert system's knowledge base for some new process will be created by assembling causal and fault models of the process units from a library that was created in the past. Then by driving this deep-level model-based diagnostic expert system by a process simulator that simulates the process behavior under different fault conditions, the system can recognize important patterns and compile them away as heuristics. For such an acquisition of compiled knowledge, learning strategies such as the failure-driven learning technique proposed in this paper would prove to be useful.

In this paper, we have proposed a failure-driven learning scheme based on causality as a means of automatically acquiring compiled knowledge and refining existing heuristics. This technique refines a heuristic in the event that the hypothesized causal origin offered by the heuristic is incorrect. In this manner, the heuristic is strengthened, thereby preventing it from firing if a similar situation were to arise. Using its deep-level knowledge of the process, the expert system draws inductive inferences from causal models to determine why the hypothesis proposed by the heuristic is inconsistent with the current state of the process and uses this knowledge to update the heuristic rule base. The learning mechanism is coupled to an existing object-oriented two-tier architecture such that new diagnostic heuristics are acquired and existing heuristics revised as the system accumulates diagnostic experience. In an effort to further enhance both diagnostic efficiency and accuracy, research continues in devising new strategies, based on the deep-level knowledge of the system, for establishing the proper strength

(specificity) or weakness (generality) of the new and updated heuristics that result from knowledge base revisions.

Acknowledgment

This work was supported in part by Grant No. 1 R01 OH02352-01A1 from the National Institute for Occupational Safety and Health of the Centers for Disease Control. The authors also acknowledge the support of the International Business Machines Corporation through the Aurora Project.

Literature Cited

- Himmelblau, D. M., *Fault Detection and Diagnosis in Chemical and Petrochemical Processes*, Elsevier (1978).
- Michalski, R. S., "A Theory and Methodology of Inductive Learning," *Machine Learning*, R. S. Michalski, ed., Tioga, Palo Alto, CA, 83 (1983).
- Mitchell, T. M., "Generalization as Search," *Artific. Intell.*, **18**, 203 (1982).
- Mitchell, T. M., P. E. Utgoff, and R. Banerji, "Learning by Experimentation: Acquiring and Refining Problem-Solving Heuristics," *Machine Learning*, R. S. Michalski, ed., Tioga, Palo Alto, CA, 163 (1983).
- Pazzani, M. J., "Refining the Knowledge Base of a Diagnostic Expert System: An Application of Failure-Driven Learning," *Proc. Nat. Conf. Artific. Intell.*, Philadelphia, 1029 (1986).
- Politakis, P., and S. M. Weiss, "Using Empirical Analysis to Refine Expert System Knowledge Bases," *Artific. Intell.*, **22**, 23 (1984).
- Reiter, R., "A Theory of Diagnosis from First Principles," *Artific. Intell.*, **32**(1), 57 (1987).
- Rich, S. H., and V. Venkatasubramanian, "Model-Based Reasoning in Diagnostic Expert Systems for Chemical Process Plants," *Comput. Chem. Eng.*, **11**(2), 111 (1987).
- Steels, L., and W. Van de Velde, "Learning in Second-Generation Expert Systems," *Knowledge-Based Problem Solving*, Prentice-Hall, Englewood Cliffs, NJ, 270 (1985).
- Venkatasubramanian, V., and S. H. Rich, "An Object-Oriented Two-Tier Architecture for Integrating Compiled and Deep-Level Knowledge for Process Diagnosis," *Comput. Chem. Eng.*, **12**(9/10), 903 (1988).

Manuscript received Dec. 15, 1987, and revision received Nov. 11, 1988.