Volume 08 Issue 04 Published, November 17, 2004 ISSN 1535-864X

Intel® Technology Journal

Toward The Proactive Enterprise

Bayes Network "Smart" Diagnostics

A compiled version of all papers from this issue of the Intel Technology Journal can be found at:
http://developer.intel.com/technology/itj/index.htm

Bayes Network "Smart" Diagnostics

John Mark Agosta, Corporate Technology Group, Intel Corporation
Thomas Gardos, Information Services and Technology Group, Intel Corporation

Index words: diagnostics, troubleshooting, fault analysis, Bayes networks, knowledge engineering

ABSTRACT

Formal diagnostic methods are emerging from the machine-learning research community and beginning to find application in Intel. In this paper we give an overview of these methods and the potential they show for improving diagnostic procedures in operational environments. We present an historical overview of Bayes networks and discuss how they can be applied to diagnosis. We then give an illustration of how they can model the faults in a vacuum subsystem of a manufacturing tool.

INTRODUCTION

Substantial portions of Information Technology (IT) and manufacturing operations budgets are dedicated to diagnostics. In most enterprises today diagnostic methods are generally ad-hoc and rely on undocumented knowledge of a few experts. These practices commonly result in excessive down time of critical infrastructures and in wasted expense associated with unnecessary tests and occasionally unnecessary repairs. This situation will continue as manufacturing and IT operations become more automated, thereby reducing the need for human operators, while at the same time increasing the amount of equipment critical to keeping the business running. For example, in the next generation of semiconductor fabrication plants, most of the wafer handling will be automated, leaving maintenance as the predominant operations cost. The problem is complicated by the fact that the most highly skilled diagnostic staff are usually the most sought after for new deployment projects, taxing the availability of diagnostic expertise for on-going operations.

Bayes Networks

As it turns out, rigorous methods are emerging from the machine-learning research community that begin to address these problems. One of the most promising is Bayes networks, a formalism based on probability and graph theory. (Bayes networks are also known as "belief networks," "Bayesian networks," or "causal networks.") Formally, a Bayes network models the probability distribution of a set of random variables as nodes in a graph, and it models the probabilistic dependencies among variables as arcs in the graph. Causal relationships are represented by conditional probabilities, encoded in the distributions associated with each node, to capture the strength of the relationship. Such directed graphs visualize the probabilistic relationships among variables. Most important is that Bayes networks model independence among variables; the network contains only a small fraction of the possible dependencies and this makes working with them manageable. Just as the network graph makes the model design easy to grasp, it also makes possible efficient computation of probability updates.

Applied to diagnostic modeling, the network's random variables represent events, divided roughly into two classes: evidence (e.g., tests or symptoms) that can be directly observed, and root causes (e.g., component failures) that cannot be observed and must be inferred from evidence. We treat diagnostic models as causal Bayes networks. Cause-effect relationships beginning with root problems and the tests or symptoms that result are both elicited from experts and learned statistically with support of operational data.

Once the model is developed, diagnosis proceeds by inverting the conditional dependence relationships applying a generalization of Bayes rule, and thus calculating the probability of root causes conditioned on evidence provided by the user. The result is a ranked list of most likely root causes based on the evidence provided so far and a second ranked list indicating the most discriminating diagnostic tests to perform next. This methodology shows some very interesting and valuable properties:

- The ability to diagnose multiple simultaneous root causes, the combination of which may have never been anticipated by the experts.
- The dynamic creation of the diagnostic sequence–effectively, the flow chart.

- The ability to submit and retract evidence and recompute its effect in any order during the diagnostic process.

Bayes Networks as a Knowledge Engineering Tool

Another valuable result of this process is that the expert knowledge is captured and thus the diagnostic model forms the basis of a knowledge engineering tool. Moreover, the expert knowledge is encoded in a manner such that diagnostic inference can then be computed–a capability that most knowledge management approaches cannot deliver.

The information that comprises the Bayes network exists in an informal and unstructured way in equipment maintenance logs, but most of it never gets applied to subsequent occurrences. The Bayes network is a concise way to encode diagnostic knowledge from a combination of engineering knowledge and statistical data in such a way that all possible test and observation diagnostic sequences can be generated from it. By their nature, breakdown events are rare, and a troubleshooting model cannot be built solely by a data-driven approach.

Bayes network diagnostics are used in several commercial implementations, most notably by Microsoft, in General Electric's Condition Forecaster* tool and for troubleshooting websites (see, for example, "Parts America.com" [19]).

Intel has sponsored an annual Bayesian Application Workshop in cooperation with the Uncertainty in Artificial Intelligence conference where commercial implementation projects are presented [4].

PROBABILISTIC LOGIC AND BAYES NETWORKS

Why is probability the right way to reason about diagnosis? The reasons may appear obvious to the reader, but the topic has been the source of much discussion, both among students of logic and of statistics. This section offers a short answer to the question. A reader uninterested in the question may want to skip directly to the example in the next section.

There is a rich history of theory that Bayes network diagnostics draws upon, as the name "Bayes" suggests. Thomas Bayes was an 18th-century cleric whose name is associated with a comprehensive concept of probability, i.e., that any uncertainty can be represented as such. The Bayesian approach weaves together statistics, economics, and machine learning. It is the starting point from which the theory of rational decision-making has been developed, and the one on which this approach to diagnostics is based. A rational solution incorporates all the decision maker knows, to the point where the decision can stand in for the judgment of the decision maker. This is a tall order, but it offers a first principles approach to solving a problem. A Bayesian solution presents the full rationale for the decision, which among other things, gives a justification for automating the decision.

Bayes network diagnostics applies a decision-theoretic concept of probability as the basis to measure information. The cost of achieving a successful diagnostic outcome is a function of the time and cost spent collecting information, and this is a significant part of modeling diagnosis. The "value of information" (VOI) computation discussed in the section "The Value of Diagnostic Observations" minimizes the diagnostic steps to isolate faults. This is a good approximation to minimizing the cost to achieve the correct diagnosis, and this can easily be improved upon by assigning different costs to performing tests. The full decision optimization problem can be addressed by embedding the diagnostic Bayes network in an optimization framework, but this is not typically justified [10]. Interestingly, see Bayer-Zubek [5] for a novel Bayesian approach that does not take advantage of a Bayes network, but does compute an optimal diagnostic policy.

Cox's Theorem as the Foundation for Probabilistic Logic

Probability is a means to express and reason with degrees of uncertainty precisely. This section outlines a derivation for probability, and the rules for updating probabilities, that is more widely applicable than conventional statistics-based derivations, but maintains consistency with them. The conclusions presented here are based on an argument first presented by R.T.Cox [7, 13, 18] He begins with three premises:

I. **The uncertainty of an event is described by a real number**. It is possible to find an equivalence between any real, uncertain value and a *certain equivalent*. For instance, you may offer a price for a used car before having the opportunity to inspect it. Your offered price weighs the possible different conditions of the car that are uncertain; but your offered price is a certain quantity, called the *expected value* of the uncertain quantity, the value of the car. A *probability* is just the expected value of an event outcome, where the certain occurrence of

* Other brands and names are the property of their respective owners.

the event has value *1* and its negation, the absence of the event, has value *0.* In most expositions the concept of expectation is derived from probability. In this case, similarly to Whittle [24] the reverse is more natural.

II. **Probabilities combine in common sense ways.** When events are certain, probabilistic logic is reduced to familiar Aristotelian logic. When probabilities can be defined by frequencies of event occurrences, probabilistic logic is consistent with counting frequencies.

If events are not certain, then to combine them and obtain their joint probability, we need to keep track of the *state of information* of each event, say *A* and *B*, in order to combine them. This is represented by another event or proposition, *C*, called the *conditioning* event. The probability of the *conditioned* event (e.g., *A, B*) is supported by *C.* Conditioning is shown by a vertical bar "|", hence "*A* given *C*" is written $(A \mid C)$. Unlike the Boolean algebra of certain propositions, the probabilities of uncertain propositions must have consistent conditioning to be combined, and the outcome of their combination will depend strongly on their conditioning. An essential difference when working with uncertainty is that uncertain inferences are "non-monotonic" in the inclusion of changing conditioning by additional evidence; additional evidence may weaken an inference, whereas once a logical proposition is proven, no additional consistent fact will change the conclusion.

III. **Combining the same knowledge in different orders gives equivalent results.** The largest part of reasoning with probabilities is reasoning about how new information–a new fact or assumption–changes our belief in a proposition, as represented by a probability. This is called *belief update.* Solve this, and the method for reasoning with probability is completed.

The specific problem of belief update can be stated as deriving the combining function, *f(),* for the probabilities *A* and *B,* where one is conditioned on the other, and they share a common state of information *C*:

$$pr(A \text{ and } B \mid C) = f(\, pr(A \mid B \text{ and } C),\, pr(B \mid C)\,)$$

To make a long story short, the derived combining rule, *f(),* not surprisingly turns out to be multiplication. Perhaps what is surprising is that Cox's theorem derives the conventional probability rules without recourse to making assumptions about sample spaces, event frequencies, set theory, additivity of the probability of mutually exclusive events, and the definition of conditional probability measures. **Thus we have a general result about reasoning under uncertainty that is consistent with, and subsumes conventional notions of statistical probability.** From this "information update" combining rule it is an algebraic derivation to obtain the update rule known as the Bayes rule:

$$pr(A \mid B \text{ and } C) \propto pr(B \mid A \text{ and } C)\, pr(A \mid C)$$

Reasoning About Causes

The notion of cause is closely entwined with probability. A cause may function unreliably, but the cause itself–as a general principle–is not at question. For example, we don't question the tendency of plumbing to wear out and leak with time, but that is not saying that any specific valves or pumps will do so. The inverse question, of reasoning from specific examples to establishing the general cause is the problem of inference. Statistical inference is applied to prove the existence of a cause in principle. In comparison, in working on diagnostic problems our models will select from a set of general causes which specific cause or causes can explain a breakdown.

Why not use an entirely qualitative (e.g., symbolic) calculus? Because there aren't any that don't reduce to computing with probability numbers [14], and because there is a bonus in using probability-based logic–it can be extended to statistical techniques. This has a direct benefit when data are available from which to learn Bayes network probabilities. Bayesian statistics consider the proper way to combine expert judgment and statistical knowledge. For a comprehensive review see especially the article by Heckerman in [16].

The models we develop work fine with the rank-ordered probability numbers estimated by the technicians. The arguments why Bayes net methods work well with order of magnitude probabilities and other methods don't are discussed extensively in the literature [21].

History of Bayes Network Diagnostics

Early work in perceptrons, precursors to neural networks, was sympathetic to Bayesian approaches. Progress in Bayesian approaches to automated reasoning along with other quantitative reasoning methods stagnated when purely symbolic approaches became popular in the early 1980s. The current blossoming of Bayesian methods coincides with Pearl's introduction of "Belief networks" [20] and Heckerman and others work on the Pathfinder project [9], as an improvement on the Mycin medical diagnostic rule-based expert system.

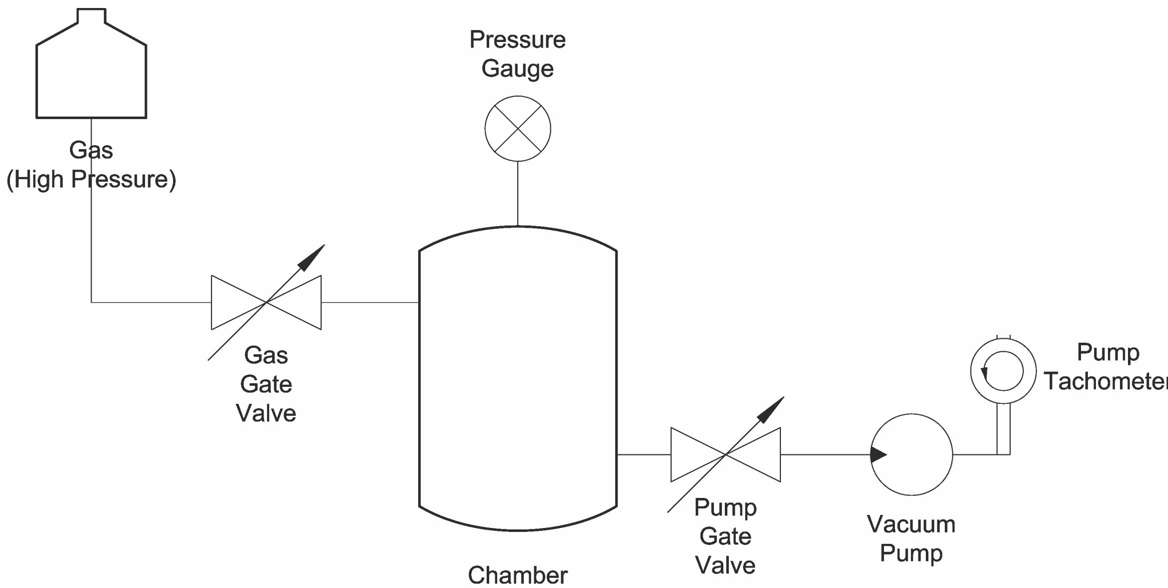

Figure 1: Chamber, valve, and pump assembly for diagnostic example

Several members from this group were early members of Microsoft Research, and they implemented a wide variety of applications for computer and software configuration diagnostics. About the same time, the field was spurred on by general-purpose Bayes networks solution methods by "join-tree" propagation, published by Lauritzen and Spiegelhalter [18]. There is also a significant Bayes network component in machine-learning research that has contributed to the growth of the field.

AN EXAMPLE DIAGNOSTIC PROBLEM

Consider the following system shown in Figure 1 that is loosely based on a chemical-vapor deposition process in silicon wafer manufacturing [1][2]. In this simplified system the chamber is pumped empty of all the reacted gas so that it is clean and ready for the next process to take place. The gas gate valve is closed to maintain pressure. The pump gate valve opens and the vacuum pump removes the reacted gas and pumps down the chamber to a low vacuum pressure of about $1/100^{th}$ of an atmosphere[1]. A gauge indicates the pressure inside the chamber and a tachometer indicates the rotational speed of the vacuum pump.

For this example, there are five components whose failure we may have to infer in a diagnostic process. The five components along with their operational states are as follows:

- Gas gate valve (good, leaking)
- Pump gate valve (good, leaking)
- Vacuum pump (good, insufficient vacuum)
- Pump tachometer (good, out of calibration, no reading)
- Pressure gauge (good, out of calibration, no reading)

Each component will be represented by a node in the Bayes net diagnostic model.

Assume there is a chamber pressure alarm that will sound if the chamber pressure strays outside of specified limits in the clean cycle.

Next, we characterize how a component failure may manifest itself. This is knowledge that an expert with the tool would possess.

- A failure of the gas gate valve would prohibit the chamber pressure from reaching the cleaning pressure level and would cause the chamber pressure to rise quickly after being pumped down.
- A failure of the pump gate valve would cause the chamber pressure to rise quickly after being pumped down, although the chamber would initially reach the target clean cycle pressure.
- If the pressure gauge failed or went out of calibration, it would fail to indicate that the pressure in the chamber reached its target and would not indicate accurate pressure levels for any diagnostic

[1] An atmosphere is a unit of pressure equal to the pressure of the air at sea level.

tests. This could confound the chamber clean pressure alarm and the rate-of-rise test.

- If the vacuum pump didn't operate correctly, the chamber pressure would not reach its target pressure level for the clean cycle.
- If the pump tachometer failed or went out of calibration, it would not give an accurate reading of the pump speed which would confound the vacuum pump speed reading.

The first four component failures would manifest themselves as a chamber clean pressure alarm. The last component failure could confound a diagnostic procedure but would not trigger a pressure alarm.

The following are the diagnostic procedures available to the technician. The implications of each procedure correspond to an arc from a failure node to a diagnostic test:

- *Rate-of-rise test*. This test measures whether the chamber can pump down to target pressure and measures how quickly the pressure rises with both gate valves closed. There are three outcomes: normal; fail-to-pump-down, and failed-rise. Fail-to-pump-down implicates a fault in the vacuum pump. Failed-rise implicates a fault in either of the gate valves.
- *Gas gate valve physical leak check*. This checks to see if there is a physical leak by using a handheld leak sensor. Possible outcomes are pass or fail. This test implicates a fault in the gas gate valve.
- *Pressure gauge check*. The pressure gauge is removed and placed on a test card and tested for operation and calibration. Possible outcomes are normal, out-of-calibration, and bad. This test implicates a fault in the pressure gauge.
- *Vacuum pump speed check*. The vacuum pump is set to a predetermined speed that is checked via the pump tachometer. Possible outcomes are pass and fail. This test implicates a fault in the pump or the tachometer.
- *Pump assembly leak test*. The pump and pump gate valves are checked for leaks. Possible outcomes are leak and no-leak. This test implicates a fault in the pump gate valve or pump gaskets.

Diagnostic Problem

Now suppose that an alarm sounds indicating that the chamber pressure did not reach the target vacuum pressure. The chamber pressure alarm is the primary indication that something is wrong, and a diagnostic procedure must commence. The technician must then decide the sequence of the diagnostic and repair steps, in order to get the system operational as soon as possible.

There are 15 combinations of failure for the first four components for each of which the pump tachometer may or may not fail, resulting in thirty total component failure combinations.

Once a breakdown has been identified we need to sequence the steps to isolate equipment faults; this needs to be done in a principled way to eliminate the typical uninformed "replace and test" maintenance procedures.

BAYES NETWORKS APPLIED TO DIAGNOSIS

The Diagnostic Process

As shown by the example, diagnosis starts with a general indication that something has failed–the *primary indication*–and narrows down the cause of the indication by a series of observation and test steps. The diagnostic process ends by finding one or more root causes that explain the failure.

Diagnosis as a Bayes Network

In diagnostic models the nodes fall into two sets: root causes or faults, and observations or tests. In the course of diagnosis a (hopefully small) subset of the observation variables will become certain. The subset of likely faults will be inferred from these. Faults won't be certain, except perhaps when a faulty component is replaced and its condition is observed directly.

Fault states predict observations. From observation states we can infer faults: we can see this with a two-node Bayes net (Figure 2). The arrow from fault node to test node respects the causal direction from fault to observation.

A Two Node Diagnostic Model

The simplest example of a diagnostic model has a fault variable that conditions a test variable (Figure 2).

Figure 2 Simple two-node Bayes net[2]

[2] All Bayes net visualizations in this document were created using the GeNIe software developed at the Decisions Systems Laboratory at the University of Pittsburgh and is freely available at http://www.sis.pitt.edu/~genie

Table 1: Example prior and conditional probability tables for a fault node and test node, respectively

fault	
broken	0.1
working	0.9

test	broken	working
pr(abnormal | fault) =	0.99	0.1
pr(normal | fault) =	0.01	0.9

Table 1 shows the probability assumptions for the two nodes. The fault node failure probability is 0.1 prior to the test. The probability that the test indicates abnormal given the fault node is broken is 0.99 and is referred to as the test sensitivity. The probability that the test indicates normal given the fault node is working is 0.9 and is called the specificity.

With the data in Table 1 we can calculate the marginal probability that the test result is abnormal by adding the two weighted conditional states:

$$pr(abnormal)$$
$$= pr(abnormal \mid fault)pr(fault)$$
$$+ pr(abnormal \mid working)pr(working)$$
$$= (.99)(.1) + (.1)(.9) = 0.189$$

It is largely because of the poor specificity that the test is predicted to be abnormal with probability 0.189. Figure 3 shows the nodes with probabilities indicated.

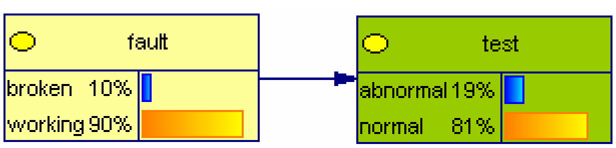

Figure 3: A two-node Bayes net

When evidence for the test is applied, the test value becomes certain, and the updated probabilities of other nodes (just one in this case) are conditioned on the evidence. Applying Bayes rule to this obtains the probability of the fault given the test, in this case for an abnormal test result (Figure 4).

Given this evidence, the probability that the fault node is broken (0.52) is now slightly greater than the probability it is working (0.48).

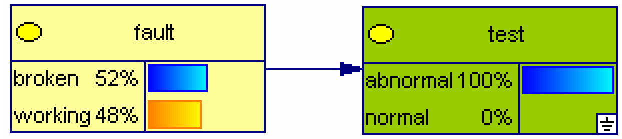

Figure 4: Inference in two-node Bayes net

Vacuum Chamber Diagnostic Model

We can now model the faults and the tests of our pressure chamber example as in Figure 5. We have color coded the nodes for easier interpretation. The five yellow nodes forming a column on the left side represent the components whose failures can be inferred. The blue node by itself in the center represents an internal unobservable state, in this case the chamber pressure. The green nodes on the right of the graph represent the diagnostic observations and tests that an operator could perform.

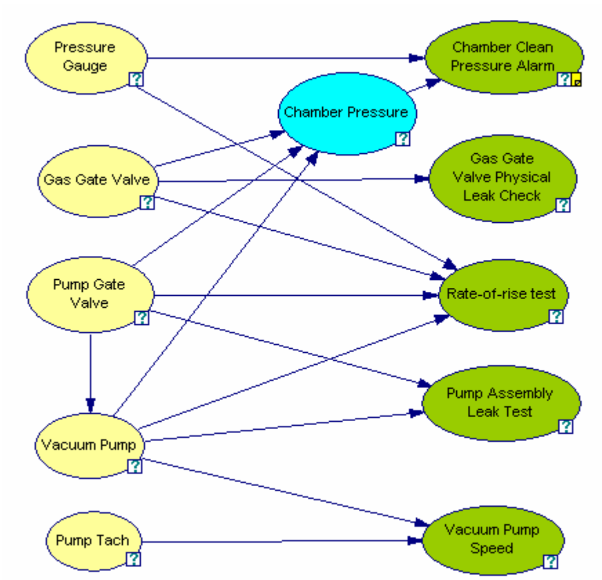

Figure 5: Bayes net diagnostic model of pressure chamber example

Figure 6 shows the same graph but with bar charts representing the marginal probabilities of the model, computed with no evidence set.

All but one of the fault nodes along the left side of the graph have no arrows entering from so-called parent nodes. The probability assignment for these nodes represents the prior probability of failure given nothing is known about the results of tests. Their values reflect the components' reliabilities. So the pressure gauge is the least likely to fail with a 1% chance, and the pump gate valve is the most likely to fail with a 6% chance.

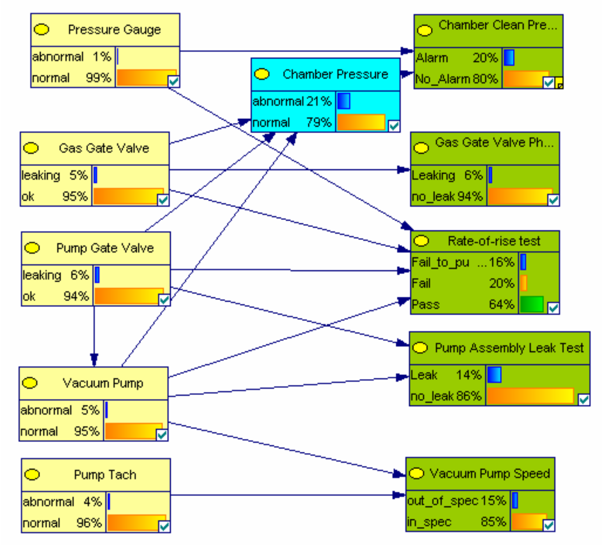

Figure 6: Example Bayes net showing marginal probabilities

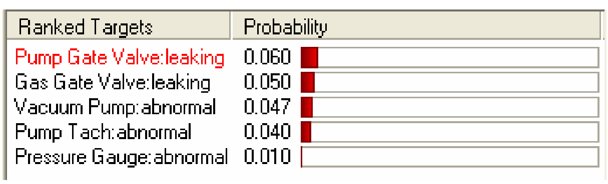

Figure 7: Ranking of possible component failures given no diagnostic test information is known

The vacuum pump node has the pump gate valve as a parent node reflecting the fact that if the pump gate valve is leaking it can make the pump more likely to fail.

Since the vacuum pump has a parent node, a conditional probability table is defined similar to the one used in Table 1. In this example, we use the values shown in Table 2.

Table 2: Conditional probability table for the vacuum pump node. The top row refers to the states of the pump gate valve which is a parent node. The abnormal and normal states refer to the vacuum pump.

Pump Gate Valve	leaking	ok
pr(abnormal | Pump Gate Valve) =	0.15	0.04
pr(normal | Pump Gate Valve) =	0.85	.96

With these data and the prior probabilities for the pump gate valve, we can calculate the marginal probability that the vacuum pump is normal by adding the two weighted conditional states as we did in the two node example to get

$pr(normal) = 0.9534.$

Since the abnormal and normal states are mutually exclusive and exhaustive, the probability of the vacuum pump being abnormal is simply the complement:

$p(abnormal) = 1 - p(normal) = 0.0466.$

We now have the probabilities of failure for all the components before observing any diagnostic tests. The ranked faults can then be displayed as a Pareto chart as shown in Figure 7. In the next section we look at how these probabilities change as we provide diagnostic evidence and update the model.

First we look at another challenge of modeling the local conditional probability for a node. Consider the chamber pressure node in Figure 5. Its possible states are normal and abnormal. Because each of its parents can take one of two states, there are eight combinations of parent states and thus eight columns in the conditional probability table that would need to be assigned (Figure 8). In general, the probability model scales exponentially with the number of variables. Most people find this level of specification unintuitive. Fortunately, if we make the reasonable assumption that the parent nodes are *causally independent* [10] given the current node, we can employ a local model, such as the Noisy MAX, that scales linearly with the number of parent variables. In short, causal independence assumes a combining rule so that it is sufficient to consider the effect separately of each parent on the child node [20].

Using a Noisy MAX model, the specification of conditional probabilities for our example is greatly simplified. Figure 9 shows the same node as the Noisy MAX model. The LEAK column on the right-hand side allows for the probability, usually small, that the node can take the abnormal state by some other means not represented in the model.

Hard Diagnostic Problems

Diagnosis can be hard for several reasons. The pressure chamber model has most of these characteristics:

- First, the primary indication may tell little or nothing about the source of the problem. An "idiot light" is an example. In the model, the chamber pressure alarm could be the consequence of four out of five faults.

Gas Gate Valve	leaking				ok			
Pump Gate Valve	leaking		ok		leaking		ok	
Vacuum Pump	abnormal	normal	abnormal	normal	abnormal	normal	abnormal	normal
abnormal	0.99685	0.991	0.9685	0.91	0.9685	0.91	0.685	0.1
normal	0.00315	0.009	0.0315	0.09	0.0315	0.09	0.315	0.9

Figure 8: Conditional probability table (CPT) for a node with three parents

Parent	Gas Gate Valve		Pump Gate Valve		Vacuum Pump		LEAK
State	leaking	ok	leaking	ok	abnormal	normal	
abnormal	0.9	0	0.9	0	0.65	0	0.1
normal	0.1	1	0.1	1	0.35	1	0.9

Figure 9: Using the Noisy MAX local probability model the conditional probabilities can be simplified for this case of a node with three parents

- The things observed are not the things that go wrong. We use the terms *observations*, *symptoms,* and *tests* for the things that can be observed, and *causes* or *faults (*or *diseases* in clinical diagnosis) for the things the diagnosis attempts to identify. Causes are *inferred* from observations. Occasionally faults can be observed directly. In this case the problem is reduced to searching efficiently and inference is not necessary. The model in the chamber pressure example is an instance where none of the failures can be observed directly.

- One root cause may cause many observations, and one observation may be non-specific. If there is a specific, accurate test for each failure and no opportunity for confusion among them (i.e., silver bullet tests), then again, diagnosis is reduced to search. In this model the “gas gate valve physical leak check” is specific to the “gas gate valve” leaking. Otherwise all faults must be isolated by a combination of observations and tests.

- Causes may cascade, so the problem is compounded by having to find the first cause that started the chain. Note how the arc from “pump gate valve leak” to “vacuum pump” models how over time the former can cause the latter to degrade.

- Tests and observations may either interfere with, or improve another test’s diagnostic value. A destructive test is an extreme example of a test that interferes with others. Tests that are only valuable in combination improve each other’s value. Interactions among test results can be modeled by arcs between tests.

- There may be numerous, possibly expensive tests that relate to numerous faults, so tests must be chosen efficiently to make progress with the diagnosis. While the size of the model is linear in the number of tests, the number of test sequences is exponential in the number of tests.

- The diagnostic problem may not be completely resolvable with the available tests, time, and money. In this situation narrowing the list of suspect faults is the best that can be done. Probability updates have the desirable property of providing useful guidance even if a complete resolution of the diagnosis is not possible.

These last three characteristics are not illustrated in this small model. However, as we see, each of these characteristics can be managed to advantage with the available modeling techniques.

In the next sections we talk more about computing inference with this diagnostic model.

Efficiently Isolating Faults by Ranking of Root Causes

As we discussed, the value of a diagnostic tool depends on how rapidly it isolates the correct fault, since even simple approaches, let alone other automated methods for diagnosis, will eventually come to a resolution of the problem. In the worst case a machinery problem can be approached by brute force replacement of all components. Clearly this is excessive in time and cost. The problem is to minimize the collection of evidence. In the best circumstances, observation and test selection will be parsimonious, and most observations will not be made.

A useful diagnostic model comprises a large number, typically hundreds, of possible observations, alarms, measurements, or tests that can be performed. These may entail costs, take time, or require materials, equipment, or skilled labor. The common thread is that it is uneconomical to seek all inputs. Therefore, there must be a tradeoff between the "diagnostic" value gained as progress toward isolating system faults and the cost expended in finding the faults We use the term "cost" to include any resource expended to generate an input to the model. In the diagnostic procedures that we consider, time will be "of the essence" and cost will be measured as a function of time to find the fault.

During diagnosis the Bayes network updates the probability of each fault based on the values of observations made to that point. Typically this is displayed as a ranking of the fault variable probabilities as shown in Figure 6. Isolating the one or more faults at the root of the problem consists of driving the probability of the faults to extremes–the probability of true faults to one and the rest to zero. An appropriate way to measure this is to compute the entropy over the set of faults [3]. Thus the most efficient set of observations is the set that maximizes the decrease in entropy of the fault set. Since the Bayes network can be used to compute the full joint probability distribution of the faults and observations under any state of information, it can be used to compute these entropy changes. The computation of these values to determine the best set of observations to make is loosely referred to as "value of information." Strictly, value of information refers to the increase in expected value of the outcomes under the optimal decision were the information value available when the decision was made [12].

In our example, if we set diagnostic evidence so that the "rate-of-rise" test is set to "fail to pump down" we get the recalculated ranked probabilities shown in Figure 10. Note that the vacuum pump probability went from 0.047 with no evidence set to 0.278 with this diagnostic test.

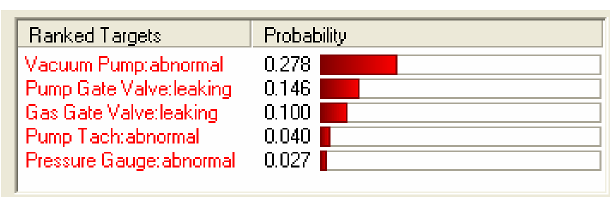

Figure 10: Ranked probabilities of component failure with rate-of-rise test set to "fail to pump down" state

If we further set a second diagnostic test, vacuum pump speed, to its "out of spec" state, we see that the vacuum pump is implicated further, increasing the probability to 0.779 (Figure 11). Note that the pump tachometer probability of failure went from 0.04 to 0.112 with the evidence of pump abnormality, representing the fact that the pump diagnosis could be confounded by a bad tachometer. The probability of the tachometer being abnormal could have been even higher except for the fact that the chamber pressure failed to pump down to vacuum during the rate-of-rise test making a pump abnormality more feasible. We would see this more clearly if we were to clear the rate-of-rise test result and only have evidence of the vacuum pump being out of spec. In that case the pump tachometer probability would jump to 0.265.

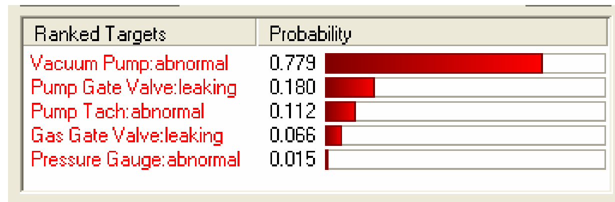

Figure 11: Ranked probabilities of component failure with both rate-of-rise test set to "fail to pump down" and vacuum pump speed set to "out of spec"

Value of Diagnostic Observations: Finding the Best Test Sequence

The challenge in diagnostic reasoning is computing which is the best observation to select next, to make progress towards finding the faults [23]. This is called *differential diagnosis*, at the stage in the diagnosis when the current information suggests several candidate faults with a high probability, to be confirmed or refuted to find what the actual faults are. As explained, a decrease in the entropy of this set of faults is the heuristic to measure progress toward confirming the faults.

The difference between the fault set entropy conditional on a set of observations and the entropy lacking the observations is called the "mutual information" between the fault and observation sets. Mutual information captures roughly the dependence between the fault and observation sets; the more positive the mutual information, the stronger the dependence. If the probability distributions of faults and observations are independent, then their mutual information is zero. Obviously such observations would be useless.

Finding the best set of observations to distinguish among the current fault candidates by mutual information becomes a hard computational problem. Each set requires computing over the Bayes network, which typically is expensive. Furthermore, mutual information over a set of observations may not be well approximated by the mutual information of the individual observations; for example, an observation that interferes with another, or is needed in combination with another, will affect the other's mutual information. In the interest of computational efficiency, in practice, these dependencies may be disregarded [3].

Finally, mutual information of any set of observations will change as other observations are made. This may be due to dependencies among observations. Also the current fault probabilities play a role: faults with extreme values are less likely to change and hence they affect mutual information less. Thus it is insufficient to rely on mutual information computed at the start of the diagnosis. The observation selection task in diagnosis may be thought of as a dynamic feature selection problem, conditional on the current observation values.

Since diagnostic value as indicated by mutual information for most of these inputs depends heavily on the current state of the diagnosis, it is worthwhile to re-compute mutual information after each observation. The procedure is thus to re-compute each observation's diagnostic value individually after each observation is taken, and the state of belief is updated, then to select the next observation with the highest value. This sequence of steps is repeated until either the probability of one or more faults approaches one, or the value of all remaining observations approaches zero. This approximate guided search method is called *myopic* test selection. Simulations in typical diagnostic models have shown that myopic test selection closely approximates optimal [11]. The mutual information computation over all possible observations as required by myopic test selection may be onerous, and further approximations may be necessary [13].

Continuing with our example, Figure 12 shows the diagnostic value of each of the tests in the chamber pressure model. Note that the rate-of-rise test shows the highest diagnostic value. If we set evidence for that test to "fail to pump down" as we did earlier in our example, that test is removed from the list. The re-ranked list is shown in Figure 13. As one would expect, the failure of the chamber to pump down implicates the vacuum pump and so the vacuum pump speed test is now highest on the list.

Ranked Observations	Diagnostic Value
Rate-of-rise test	0.286
Pump Assembly Leak Test	0.260
Vacuum Pump Speed	0.229
Chamber Clean Pressure Alarm	0.228
Gas Gate Valve Physical Leak Check	0.181

Figure 12: Diagnostic value of information for chamber pressure example before any diagnostic evidence is set

Ranked Observations	Diagnostic Value
Vacuum Pump Speed	0.240
Pump Assembly Leak Test	0.228
Chamber Clean Pressure Alarm	0.144
Gas Gate Valve Physical Leak Check	0.127

Figure 13: Diagnostic value of information for chamber pressure example after rate of rise test is set to fail to pump down

ON DEPLOYING A BAYES NETWORK-BASED DIAGNOSTIC MODEL

In our work we are modeling a subsystem of a chemical vapor deposition tool. We have been working with the domain experts, in this case the tool engineer and lead technician, to elicit the Bayes net model for that subsystem. Our network consists of 65 nodes and 81 arcs. Of those, 23 represent components likely to fail, 12 represent internal unobservable states, and 25 represent diagnostic tests. We also have three nodes that we call conditioning nodes. They are not probabilistic but rather represent some known state of the system such as the number of hours of operation of the tool greater or less than 10,000 hours.

Our experience has shown that the cause-effect relationships modeled by the Bayes net graph were quickly picked up by the domain experts. In fact, after an introductory session, the equipment engineer quickly sketched out a thirty node graph. Refining the graph proceeded more slowly, however. We calculate that we averaged about three nodes per hour during subsequent sessions. During these sessions, we not only extended the topology but also defined the conditional probabilities. Relating the probabilities to the graph was not quite as intuitive an exercise, but once we started evaluating the model with test cases, the domain experts became more comfortable with that aspect as well.

Our next steps are to develop a simple user interface to the model inference engine and then pilot the tool on the factory floor where we can measure the diagnostic performance against current methods.

FUTURE CHALLENGES

The combination of current diagnostic modeling theory and practice is mature and offers a valuable set of next-generation tools to improve manufacturing and operations practices. Looking to the future, research in the field promises several ways that current Bayes networks diagnostics can be extended to offer more value.

Intel Research has sponsored a Strategic Research Project (SRP) in Statistical Computing on Bayes Networks and Graphical Models. Many of the emerging techniques discussed in this section (in addition to the algorithms needed for diagnostic inference) have been implemented in Intel's release of the open source *Probabilistic Network Library* [22].

Diagnostic modeling can be applied to process faults in the same way the example in this paper applied it to machinery faults. In semiconductor manufacturing most problems first come to light either because a problem (a "primary indication") is detected in the process or in the product itself. For example, particle counts on the wafer or electrical tests of the dice indicate that something is wrong in the process. Conceptually, a Bayes network can model the set of causes leading up to the problem and infer from measurements which cause is at fault. It can only do this, of course, if the possible causes are understood well enough to model them, and therefore the model would be limited to well-understood parts of the process. Furthermore, process diagnostic models could initiate the use of a machinery diagnostic, since the first indication of a machinery breakdown is often detected by the process.

There are extensions to Bayes networks for temporal modeling (Dynamic Bayes Networks*)* that could be used to consider how problems evolve over time. "Unscheduled maintenance" by its nature does not exploit the temporal aspect. It takes place entirely when the breakdown occurs. A diagnostic model that tracked the evolution of the machine or process state over time could predict the machine state, for use in predictive maintenance.

The dynamics of the manufacturing process itself may be amenable to Bayes network modeling. The difficulties with modeling poorly understood processes apply here also. A current trend in research in the field is to build a preliminary model from the process diagnostic model and use this as a prior distribution for learning parts of the model from data where data are available. The advantage of such an approach is the insight the diagnostic model offers into the workings of the larger model. Techniques for learning model structure are an area of active research.

There are also Bayes networks extensions to decision making, called Partially Observable Markov Decision Processes (POMDPs). Control problems that arise in Automated Process Control (APC) can be formulated and solved as POMDPs. POMDPs have the advantage of transparency, something they share with dynamic Bayes networks. POMDPs are computationally challenging and also an area of active research.

Bayes networks and their extensions have applications generally to modeling where uncertainty and optimization apply, such as automating network and server cluster management, product test sequence optimization, and supply chain management.

SUMMARY

Innovation occurs as a combination of theoretical and practical advances. Bayes networks offer a comprehensive and principled theory for diagnostic modeling. Their adoption in industry has been slow, due in part to the learning curve of the theory to apply them. Although such models prove challenging to build, they are capable of automating hard diagnostic tasks.

ACKNOWLEDGEMENTS

We acknowledge the assistance and contributions to this work by the DSL lab, the University of Pittsburgh, and Steve Zambroski, Jolyon Clarke, and Quoc Truong. All examples and network illustrations were created with DSL's Genie* software.

REFERENCES

[1] "Integrated Circuit," *Encyclopædia Britannica Premium Service*, Oct. 1, 2004, http://www.britannica.com/eb/article?tocId=34340*.

[2] "Semiconductor device fabrication," Wikipedia. 2004, http://en.wikipedia.org/wiki/Semiconductor_device_fabrication*.

[3] Agosta, J. M. and J. Weiss, "Active Fusion for Diagnosis, Guided by Mutual Information Measures," *Proceedings of the 2nd International Conference on Information Fusion,* Sunnyvale, CA, pp. 337-344, July 1999.

[4] Agosta, J. M. and O. Kipersztok, editors, "Proceedings of the 2nd Bayesian Modeling Applications Workshop" of *Uncertainty in Artificial Intelligence,* 20, Banff, CA, 7 July 2004.

[5] Bayer-Zubek, Valentina, "Learning Diagnostic Policies from Examples by Systematic Search," *Uncertainty and Artificial Intelligence, 20,* pp. 27-34, 2004.

[6] Cowell, R. G., Dawid, A. P., Lauritzen, S. L., Spiegelhalter, D. J., *Probabilistic Networks and Expert Systems*, Springer-Verlag, NY, NY, 1999.

* Other brands and names are the property of their respective owners.

[7] Cox, R.T., *The Algebra of Probable Inference,* John's Hopkins Press, Baltimore, MD, 1961.

[8] Druzdzel, M.J. and van der Gaag, L.C., "Building probabilistic networks: Where do the numbers come from?", *IEEE Transactions on Knowledge and Data Engineering 12*, pp. 481-486.

[9] Heckerman, D., B. Nathwani., "An evaluation of the diagnostic accuracy of Pathfinder," *Computer and Biomedical Research, 25*, pp. 56-74, 1992.

[10] Heckerman, D., J. S. Breese, "Decision-Theoretic Troubleshooting: A Framework for Repair and Experiment " *Microsoft Technical Report* MSR-TR-96-06, Redmond, WA: March 1996.

[11] Heckerman, D., J. S. Breese, and Koos Rommelse, "Decision Theoretic Troubleshooting," *Communications of the ACM,* Vol. 38, No. 3, pp. 49-56, March 1995.

[12] Howard, R. A., "Information Value Theory*" IEEE Transactions on Systems Science and Cybernetics,* Vol., SSC-2, No. 1, pp. 22-26, August 1966, in *Readings in Decision Analysis,* Stanford Research Institute, Menlo Park, CA, 1977.

[13] Jagt, R. M., "Support for Multiple Cause Diagnosis with Bayesian Networks," *M.S. Thesis, Information Sciences Department*, U. of Pittsburgh, 2002.

[14] Jaynes, E.T., *Probability Theory: the Logic of Science,* University Press, Cambridge, UK, 2003.

[15] Jensen, F. V., *Bayesian Networks and Decision Graphs,* Springer-Verlag, NY, NY, 2001.

[16] Jordan, M. Ed., *Learning in Graphical Models,* MIT Press, Cambridge, MA, 1999.

[17] Kalagnanam, J. and M. Henrion, "A Comparison of Decision Analysis and Expert Rules for Sequential Diagnosis," in R. D. Shachter, T. S. Levitt, L. N. Kanal and J. F. Lemmer, Eds. *Uncertainty and Artificial Intelligence, 4,* pp. 253-270, New York, NY: North Holland, 1990.

[18] Lauritzen, S. and D. Spiegelhalter, "Local computations with probabilities on graphical structures and their applications to expert systems," *J. Royal Statistical Society B*, 50, pp. 157-224, 1988.

[19] *Parts America; The Fastlane to Auto Parts and Accessories* http://www.partsamerica.com*.

[20] Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan-Kaufman, San Mateo, CA, 1988.

[21] Pradhan, M., M. Henrion, G. Provan, B. Del Favero, and K. Huang, "The sensitivity of belief networks to imprecise probabilities: An experimental investigation," *Artificial Intelligence Journal, 84:1-2,* pp. 357, July 1996.

[22] "Probabilistic Network Library" on SourceForge https://sourceforge.net/projects/openpnl/*.

[23] van der Gaag, L.C. and M. Wessels, "Selective Evidence Gathering for Diagnostic Belief Networks," *AISB Quarterly, 86,* pp. 23-34, 1993.

[24] Whittle, P., *Probability via Expectation*, Springer-Verlag, NY, NY, 1992.

AUTHORS' BIOGRAPHIES

John Mark Agosta is part of the Statistical Computing SRP in Intel Research. Prior to joining Intel, he built Bayes networks for medical, avionics, and utility and automobile clients at Knowledge Industries and at SRI International. John received his Ph.D. degree in the Engineering-Economic Systems Department (now Management Science and Engineering) of Stanford University in 1991. At Stanford he participated in the early development of Bayes networks. His e-mail is john.m.agosta at intel.com.

Thomas Gardos has been with Intel since 1993 and is currently a senior researcher in Intel's IT Research team. His main interests are multimodal interfaces for mobile computing applications, Bayesian network-based diagnostic systems and digital video coding. Tom has thirteen patents in the areas of digital video coding and signal processing. He received his Ph.D. degree in Electrical Engineering from the Georgia Institute of Technology in 1993. His e-mail is thomas.r.gardos at intel.com.

For further information visit:

developer.intel.com/technology/itj/index.htm