Paper:

R., W. & R.S., R. (1997). A semantically constrained Bayesian network for manufacturing diagnosis. *International Journal of Production Research, 35*(8), 2171-2188.

http://dx.doi.org/10.1080/002075497194796

International Journal of Production Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tprs20

A semantically constrained Bayesian network for manufacturing diagnosis

R. W. Lewis & R.S. Ransing

Published online: 15 Nov 2010.

INT. J. PROD. RES., 1997, VOL. 35, NO. 8, 2171–2187

A semantically constrained Bayesian network for manufacturing diagnosis

R. W. LEWIS†‡ and R. S. RANSING†

The diagnostic problem is posed as recognizing patterns in rejection data and the subsequent mapping to causes. A new network architecture has been proposed which should overcome many of the disadvantages of the existing diagnostic tools. The network is based on the authors' earlier work (Ransing *et al.* 1995) on representing the causal relationship in the defect-metacause-rootcause form. Although the algorithm is based on the Bayesian analysis, many of the laws of probability have been altered to suit the complexities involved. For example, the notion of conditional probability has been generalized to enable the belief revision even in the presence of partial evidence. The inherent presence of the degree of ignorance or uncertainty in the quantification of a relationship has also been considered. Rigorous constraints, again based on the laws of probability, have been developed to check the consistency among the network values. The network is required to be initialized with only a few values or the range for the same and then a set of globally consistent values is generated automatically and efficiently. Using the most suitable set of consistent values, the diagnosis is performed using the generalized Bayesian analysis. The network has been tested for a pressure die casting process, however, it is generic in nature and can also be applied to other manufacturing processes.

1. Introduction

Diagnosis has always been a prime area of research in Artificial Intelligence and related fields. New methodologies have evolved to solve such problems and older techniques have been refined to perform better. The computer programs which perform these tasks are referred to by a variety of names in the open literature e.g. expert systems, knowledge based systems, decision support systems, rule based systems, model based systems etc.

Many of the diagnostic paradigms were first developed for medical diagnosis and were later applied to engineering problems. The use of rules as knowledge representation can be traced back to the mid seventies (Davis *et al.* 1977). In those days, the program in question was simply a collection of conditional statements and the rules were deterministic in nature. Later, uncertainty was introduced into the systems by using either probabilistic models or other theories such as certainty factor (Shortliffe 1976, Patterson 1990). Currently, many commercial shells are available which do the inferencing, uncertainty handling and a user is required only to feed in rules. In the late eighties the application of this approach to manufacturing diagnostic problems was reported by Creese (1988) and Piwonka (1989). In these systems

Revision received October 1996.

† Institute of Numerical Methods in Engineering, University of Wales Swansea, Swansea, SA2 8PP, UK.

‡ To whom correspondence should be addressed.

causal knowledge was stored in the form of rules and the computer program identified causes for a single defect type.

Often a production engineer will form an opinion based on control charts (Juran 1964). In such cases, heuristics are captured in the form of rules which are coupled with rule based diagnostic systems (Ward *et al.* 1991, Kuro and Mital 1993). However, in machining processes such as turning, milling, drilling etc. the causal relationships are not as complex or highly interlinked when compared to those occurring for metal forming processes such as sand casting, die casting and injection moulding. There is a distinct feature in the causal relationships which exist for metal forming processes that, apart from the fact that they are highly interlinked, constrains the ability of rule base architecture to capture the knowledge. This may be stated as follows:

A certain combination of defects generally occur as a result of a combination of assignable causes. The reduction in the belief value in an assignable cause, due to the non-occurrence or the partial occurrence of a related defect, depends on:

(1) the relative occurrence of a defect w.r.t. to other defects and
(2) the assignable cause under consideration.

The work proposed in this paper is based on this axiom. The existence of such a highly interlinked causal relationship makes such reductions in belief values at all possible. Also, the common occurrence of two or three defects may strengthen the belief in some of the common assignable causes. The term *partial occurrence* of a defect implies that the proportion of components rejected due to this defect, among the total rejected components, is neither significantly low nor significantly high. Occurrence or non-occurrence of a defect means that the proportion of components rejected due to the defects is either significantly high or significantly low.

2. Current diagnostic approaches

The advantages and limitations of existing approaches in attempting to solve the class of diagnostic problems defined above are discussed in this section.

2.1. *Rule based systems*

The above axiom requires the association of a fraction indicating the strength of occurrence, with each symptom (or defect). The values zero and one correspond to the non-occurrence and the occurrence of a symptom, respectively. Then, a typical rule would be formed as follows:

IF Symptom 1 (s_1%) and Symptom 2 (s_2%) and Symptom 3 (s_3%)...
THEN Cause 1 (c_1%) and Cause 2 (c_2%) and Cause 5 (c_5%).
where, s_i is the strength of occurrence of the ith symptom
and c_j would be the certainty factor in the jth cause.

This approach has two major limitations.

(1) The number of rules can increase dramatically as every combination of defects, with varying degrees of strengths, requires mapping to a combination of causes.
(2) The precise articulation of each rule is extremely difficult. Even if such rules are formed, the level of confidence in each rule is debatable.

These limitations are particularly true for the manufacturing processes such as sand casting, die casting or injection moulding. However, Fujikawa and Ishi (1995) have shown the successful application of a rule based system using probabilistic analysis for the diagnosis of the forged components. The important point to note in their work is the fact that they have proposed an adaptive scheme to update the probability values. The authors also believe that some means should be developed to generate the probability values automatically. A great deal of difficulty was experienced by the authors in extracting probability values from the foundrymen.

2.2. *Probabilistic model*

This model was proposed by the authors (Ransing *et al.* 1995) during their earlier work on the development of a robust algorithm for defect analysis. An innovative way of *eliciting* knowledge was illustrated by introducing the concept of metacauses to the causal relationship. The causes are divided into two partitions, firstly, *metacauses* i.e. the scientific rationale influencing the occurrence of defects, and secondly, *rootcauses* i.e. the process, design and material parameters which influence the occurrence of metacauses.

In their work, Ransing *et al.* (1995) made the following assumptions.

(1) The prior probability distribution was defined on the *singletons* of S, M and C respectively i.e. single elements of the corresponding set. The occurrences of these singletons were not interpreted as mutually exclusive events.
(2) The singletons were considered as statistically independent.
(3) The influence of metacauses (or rootcauses) on the occurrence of defects (or metacauses) respectively, was quantified via a conditional probability.

However, in a realistic casting environment, the practical implementation of the probabilistic approach requires consideration of the following points:

- Laws of probability are deterministic e.g. $P(A) + P(\neg A) = 1$. It does not support the concept P(A) when the degree of occurrence of A is between 1 i.e. P(A) and 0 i.e. $P(\neg A)$. Similarly it can be argued for the conditional probability. Only two conditional probabilities exist. Either $P(A \mid B)$ or $P(A \mid \neg B)$. A real life situation requires concepts such as fuzzy sets to model this phenomenon.
- The influence of an assignable cause is quantified for every combination of defects. The practical drawback of this is that a very large number of probabilities must be known in advance, and the acquisition of these is difficult.
- The probabilistic model (Ransing *et al.* 1995) requires fewer probability values because the occurrence of a defect was assumed independent to the occurrence of other defects. However, this approach was constrained in its ability to evaluate the evidence which has been scaled from zero to one. For example, the proportion of defective components with a particular among the total number of defective components is be scaled from zero to one.
- The defects–metacause–rootcauses relationships are highly interlinked. Any change in the probability value propagates to other probability values, and consistency among all the values is difficult to maintain if done manually by a trial and error method.

2.3. *Fuzzy sets*

Fuzzy sets (Zadeh 1965, 1978) offer a promising role as an alternative knowledge representation scheme. The main advantage lies in the ability to categorize the membership of a set by introducing a notion of the degree of membership (from zero to one). In conventional set theory, a member is either included in a set, or excluded. The concept of membership enables us to capture the notion of partial evidences as described earlier. Also, for causes, the degree of membership can quantify the associated posterior belief. The use of fuzzy sets as a knowledge representation scheme has been discussed in detail by Pedrycz (1993).

A diagnostic model based on fuzzy sets has previously been developed by Pandelidis and Kao (1990). Even though the model has only been validated for the injection moulding process, it is generic in nature and contains the following key points.

- It is recognized that the defect-cause relationship is extremely complex. i.e. a defect can be a manifestation of many causes and a single cause can influence the occurrence of many defects. For the many defects observed a combination of causes which best explains them may be diagnosed using fuzzy sets.
- The strength of the association between a cause and defect is quantified and stored as a weight. During a diagnostic process, these weights are modified by a factor which will depend on the current process status.
- The observed defects are quantified according to their ability to reflect the relative importance of each defect with respect to others. This point strengthens the authors' premise regarding the existence of partial evidences and their significance on the diagnostic results.
- From all possible lists of causes, only a particular combination of causes will be selected which is based on a certain predefined criterion e.g. a minimum cover criterion.

However, in order to generate a more robust diagnostic model, consideration should be given in the following points.

- Only the observed defects were considered in the diagnostic process. Consideration of the fact that some defects did not occur may help during the belief revision process.
- Only three discrete categories for the severity of defects were considered. A uniform scaling of partial evidences from zero to one may improve the quality of results.
- In any manufacturing environment, the chances are that certain defects will occur more often than others. Similarly, the chances of some causes being more responsible than others should be taken into account. A consideration of these prior beliefs in any diagnostic process is of importance and hence should be accounted for.
- The precise estimation of all the weights in the linkages is a very difficult task for any single person. Even after such an estimation, opinions may differ from person to person. However, it is much easier to estimate a range of values and then automatically seek the best value. The methodology in arriving at the best choice of weights and other numerical values needs a thorough investigation.

2.4. *Need for alternative algorithms*

Hence there is clearly a need to evolve a new algorithm which would overcome these practical limitations. An attempt is made in this paper to develop a theory to estimate how strongly a set of evidence (including partial evidence) can favour the possible existence of a cause. This approach is similar to the Dempster-Shafer theory (Shafer 1979) in that the provability of a hypothesis is assessed rather than computing the probability of it being true. At the same time the theory also relies on the concept of Bayesian analysis whilst revising the prior belief in a cause in the light of a set of evidence.

In the next section a new network topology has been introduced. To avoid confusion with the variables used in the probability theory, new variables are introduced. Section 3.1 discusses their physical interpretation. It also describes how to initialize a semantically constrained Bayesian network from an initial guess. In § 4 the concept of the feasible set is introduced which will incorporate all the consistent numerical values. The processing of the input vector in the Semantically Constrained Bayesian Network, i.e. the diagnostic algorithm is detailed in § 5. After the validation of this approach in § 6, the paper is concluded in § 7.

3. A semantically constrained Bayesian network

A semantically constrained Bayesian network generalizes the simplified probabilistic model (Ransing *et al.* 1995) in the following way.

- It is possible to evaluate the partial evidence i.e. the strength of the evidence is scaled from 0 to 1.
- The notions of conditional probability are generalized.
- It accepts incomplete information regarding the probability values i.e. $P(A) + P(\neg A)$ does not necessarily equal 1. In other words it allows to account for the presence of some degree of ignorance in the quantification process of the causal relationship.
- From the approximate guess of the probabilistic values, consistent and more accurate values are generated automatically.

3.1. *Construction of the network*

Figure 1 shows the three layered knowledge elicitation scheme developed by the authors (Ransing *et al.* 1995). Readers are referred to this publication for more details on generating a defect-metacause-rootcause relationship for a manufacturing process. The semantically constrained Bayesian network is based on this network. Thus this network will have defect nodes, metacause nodes and rootcause nodes interlinked in the same way as in the corresponding defect-metacause-rootcause relationship.

3.2. *Notation with physical interpretation*

3.2.1. *Degree of activation*

The partial evidence described in the previous section is quantified by introducing the concept of degree of activation.

Definition: The strength of occurrence of a defect, metacause or rootcause is quantified by a fraction a which is defined as the degree of activation.

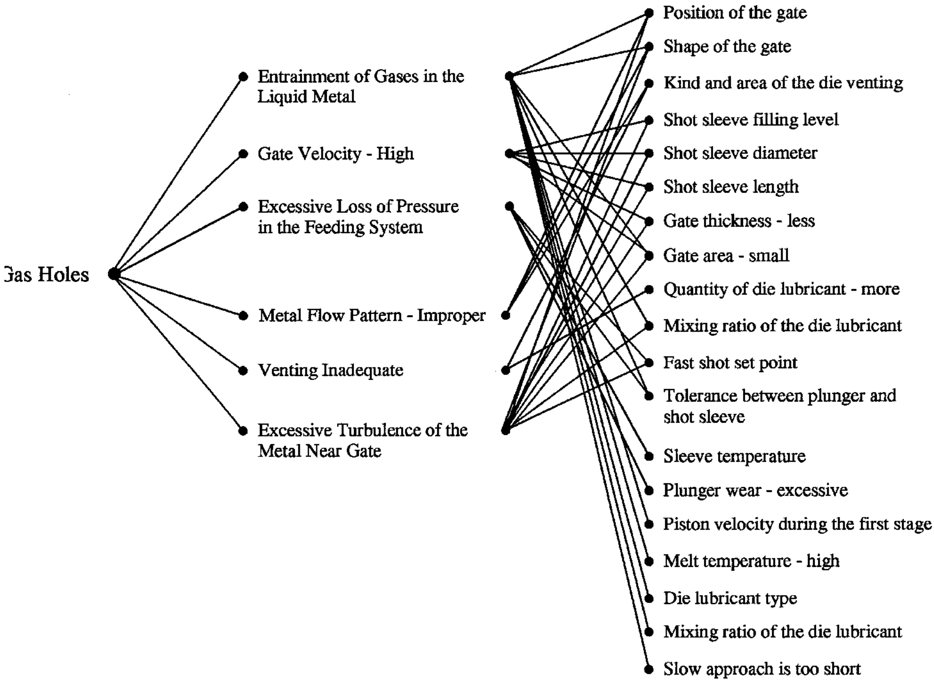

Figure 1. Causal relationship in the Defect-Metacause-Rootcause form.

Remarks:

(1) $a = 0$, indicates the non-occurrence of a corresponding defect, metacause or rootcause;
(2) $a = 1$, indicates the occurrence of a corresponding evidence defect, metacause or rootcause with maximum certainty;
(3) $a = 0{\cdot}5$, represents uncertainty regarding the occurrence of corresponding defect, metacause or rootcause. As a result, no conclusions would be drawn on the occurrence/non-occurrence of the respective defect, metacause, or rootcause.

The activation for defect nodes, input to the semantically constrained Bayesian network, is calculated from the rejection data. A small degree of preprocessing work is done to convert this data into an input vector. First, the rejection data is normalized as follows:

For the ith defect node, the normalized rejection fraction n_i is

$$n_i = \frac{\text{No. of defective components due to the defect type } i}{\text{Total number of defective components}} \tag{1}$$

The normalized rejection fractions are then mapped onto the set $[0, 1]$ to generate the input vector of degrees of activation.

$$a_i = n_i \frac{1}{\max(n_i)} \tag{2}$$

The output of the network is the degree of activation for metacause and rootcause nodes. Section 6 will detail how to generate the output from the input.

3.2.2. *Belief vectors $\boldsymbol{B}_d, \boldsymbol{B}_m$ and $\boldsymbol{B}_r$*

The degree of certainty associated in the prior estimation of beliefs for the occurrence of defects, metacauses and rootcauses are represented by the belief vectors ($\boldsymbol{B}_d, \boldsymbol{B}_m$ and $\boldsymbol{B}_r$). These differ from prior probabilities in probability theory where the probability of the occurrence and non-occurrence of an event sums to unity. This point will be considered further in the next section. However, as a result, the complementary belief vectors are required to be explicitly defined as follows.

The complementary belief vectors, $\boldsymbol{B}_d^c, \boldsymbol{B}_m^c$ and $\boldsymbol{B}_r^c$ represent the degree of certainty associated in the prior estimation of beliefs in the non-occurrence evidence and hypotheses respectively. The belief vectors for defect, metacause and rootcause nodes respectively, are represented as follows:

$$\boldsymbol{B}_d = \left[\boldsymbol{B}_{d_1}, \boldsymbol{B}_{d_2}, \boldsymbol{B}_{d_3}, \ldots, \boldsymbol{B}_{d_i}, \ldots, \boldsymbol{B}_{d_o}\right]^T \qquad \boldsymbol{B}_d^c = \left[\boldsymbol{B}_{d_1}^c, \boldsymbol{B}_{d_2}^c, \boldsymbol{B}_{d_3}^c, \ldots, \boldsymbol{B}_{d_i}^c, \ldots, B_{d_o}^c\right]^T$$

$$\boldsymbol{B}_m = \left[\boldsymbol{B}_{m1}, \boldsymbol{B}_{m2}, \boldsymbol{B}_{m3}, \ldots, \boldsymbol{B}_{m_j}, \ldots, \boldsymbol{B}_{m_p}\right]^T \qquad \boldsymbol{B}_m^c = \left[\boldsymbol{B}_{m_1}^c, \boldsymbol{B}_{m_2}^c, \boldsymbol{B}_{m_3}^c, \ldots, \boldsymbol{B}_{m_j}^c, \ldots, B_{m_p}^c\right]^T$$

$$\boldsymbol{B}_r = \left[\boldsymbol{B}_{r1}, \boldsymbol{B}_{r2}, \boldsymbol{B}_{r3}, \ldots, \boldsymbol{B}_{r_j}, \ldots, \boldsymbol{B}_{r_q}\right]^T \qquad \boldsymbol{B}_r^c = \left[\boldsymbol{B}_{r_1}^c, \boldsymbol{B}_{r_2}^c, \boldsymbol{B}_{r_3}^c, \ldots, \boldsymbol{B}_{r_k}^c, \ldots, B_{r_q}^c\right]^T$$

where o, p, q are the number of defect, metacause and rootcause nodes respectively.

Each node is associated with a threshold value (θ) which is analogous to the prior odds in a Bayesian analysis i.e. a ratio of the belief value to the complementary belief value.

$$\theta_{m_j} = \boldsymbol{B}_{m_j} / \boldsymbol{B}_{m_j}^c$$

$$\theta_{r_k} = \boldsymbol{B}_{r_k} / \boldsymbol{B}_{r_k}^c$$

3.2.3. *Initial guess for belief vectors*

The rationale used for the prior estimation of both the belief vectors $(\boldsymbol{B}_{d_i}, \boldsymbol{B}_{m_j}$ and $\boldsymbol{B}_{r_k})$ and the respective influence factors $(\boldsymbol{I}_{d_i m_j a=1}$ and $\boldsymbol{I}_{m_j r_k a=1})$ is identical to the one used in ICADA (Ransing *et al.* 1995). The prior belief in the occurrence of a defect, metacause or rootcause is arrived at by taking the ratio of the frequency of occurrence defects, metacauses or rootcauses respectively (with activation greater than 0·6) to the total number of diagnostic sessions. Note that in the case where two different defects manifest themselves for a single component, this should be marked under both defect types. Clearly, the sum of all components marked as rejected under each defect type may exceed the actual number of rejected components. This also means that the following laws of probability are not obeyed.

Laws of probability	Semantically constrained Bayesian network
$\sum_i P(d_i) = 1$	$\sum_i \boldsymbol{B}_{d_i} \neq 1$
$\sum_j P(m_j) = 1$	$\sum_j \boldsymbol{B}_{m_j} \neq 1$
$\sum_k P(r_k) = 1$	$\sum_k \boldsymbol{B}_{r_k} \neq 1$

Table 1.

The initial guess in influence factors is estimated by treating them as conditional probabilities. In these predictions, the term occurrence is broadly interpreted to cover all the cases when the respective activation is greater than 0·6. It is assumed that only 60% of the guess can be attributed to the case with activation equal to unity.

In short, $\forall i, j, k$

$\boldsymbol{B}_{d_i}$ = 0·6 (initial guess for ith defect node)
$\boldsymbol{B}_{m_j}$ = 0·6 (initial guess for jth metacause node)
$\boldsymbol{B}_{r_k}$ = 0·6 (initial guess for kth rootcause node)
$\boldsymbol{I}_{d_i m_j a=1}$ = 0·6 (initial guess for the respective defect-metacause relationship)
$\boldsymbol{I}_{m_j r_k a=1}$ = 0·6 (initial guess for the respective metacause-rootcause relationship)

In a probabilistic model, complementary belief values (e.g. $P(\neg d_i)$ get defined implicitly through the laws of probability. These laws of probability assume the existence of the complete knowledge regarding the occurrence or non-occurrence of defects, metacauses or rootcauses in a manufacturing setup. However, it has been observed by the authors that at least 50% degree of uncertainty could prevail in such precise quantification. If the same expert is asked to quantity the prior belief in the occurrence and non-occurrence of a defect or a cause separately, it is extremely likely that the sum would be less than unity. This is due to the presence of certain degree of ignorance or uncertainty in such quantifications. The Dempster-Shafer theory (Shafer 1979) also allows us to model such a degree of ignorance. Therefore, it would be wrong to apply such laws of probability for this problem. Based on the experience of the authors in this field, the following range for the summation of the belief and complementary belief values is proposed.

Laws of probability	Semantically constrained Bayesian network
$P(d_i) + P(\neg d_i) = 1$	$\boldsymbol{B}_{di} + \boldsymbol{B}^c_{di} \approx 0{\cdot}45 - 0{\cdot}5$
$P(m_j) + P(\neg m_j) = 1$	$\boldsymbol{B}_{mj} + \boldsymbol{B}^c_{mj} \approx 0{\cdot}45 - 0{\cdot}5$
$P(r_k) + P(\neg r_k) = 1$	$\boldsymbol{B}_{rk} + \boldsymbol{B}^c_{rk} \approx 0{\cdot}45 - 0{\cdot}5$

Table 2.

where, d_i, m_j and c_k represent defect, metacause and rootcauses nodes respectively.

3.2.4. *Influence factors*

Definition:

If the jth node $(\boldsymbol{B}_{m_j})$ influences the occurence of the ith node in the previous layer $(\boldsymbol{B}_{d_i})$ with activation $a = 1$, then

(a) the degree of influence of the occurrence of the jth node $(\boldsymbol{B}_{m_j})$ on the occurence of the ith node $(\boldsymbol{B}_{d_i})$ is defined by an 'influence factor', $\boldsymbol{I}_{d_i m_j a=1}$.
(b) the degree of infuence of the non-occurrence of the jth node $(\boldsymbol{B}_{m_j})$ on the occurrence of the ith node $(\boldsymbol{B}_{d_i})$ is defined by an 'adjoint influence factor', $\boldsymbol{I}'_{d_i, m_j a=1}$.

Complementary influence factors: The complementary influence factors $\boldsymbol{I}_{d_i m_j a=0}$ quantify the influence of the occurrence of the jth node ($\boldsymbol{B}_{m_j}$) on the non-occurence of the ith node ($\boldsymbol{B}_{d_i}$). Similarly, complementary adjoint influence factors $\boldsymbol{I}_{\acute{d}_i m_j a=0}$ would quantify the influence of the non-occurrence of the ith node ($\boldsymbol{B}_{d_i}$).

Remarks:

(1) Clearly, the influence factor $\boldsymbol{I}_{d_i m_j a=1}$ and the adjoint influence factor $\boldsymbol{I}_{\acute{d}_i m_j a=1}$ are a generalization of the conditional probability values in probability theory $P(E_i \mid H_j)$ and $P(E_i \mid \neg H_j)$ respectively. They are a measure of the strength of the relationship between two nodes in consecutive layers corresponding to the respective degree of activation.
(2) $\boldsymbol{I}_{d_i m_j a}$ and $\boldsymbol{I}_{\acute{d}_i m_j a} \in [0, 1]$
(3) If the occurrence and non-occurrence of the jth node $(\boldsymbol{B}_{m_j})$, does not influence the occurrence and non-occurrence of the ith node $(\boldsymbol{B}_{d_i})$ in the previous layer then, $\boldsymbol{I}_{d_i m_j a=1} = \boldsymbol{I}_{\acute{d}_i m_j a=1} = \boldsymbol{B}_{d_i}$ and $\boldsymbol{I}_{d_i m_j a=0} = \boldsymbol{I}_{\acute{d}_i m_j a=0} = \boldsymbol{B}^c_{d_i}$.
(4) The values stored in the semantically constrained Bayesian network are always was associated with a unit activation unless and otherwise specifically mentioned.
(5) The causal relationship has been characterized by four independent values viz. $\boldsymbol{I}_{d_i m_j a=1}$, $\boldsymbol{I}_{d_i m_j a=0}$, $\boldsymbol{I}_{\acute{d}_i m_j a=1}$ and $\boldsymbol{I}_{\acute{d}_i m_j a=0}$ which correspond to $P(E_i \mid H_j)$, $P(E_i \mid \neg H_j)$, $P(E_i \mid \neg H_j)$ and $P(\neg E_i \mid \neg H_j)$ respectively. However, the following law of probability is again relaxed.

These remarks constitute the following changes in the laws of probability.

Laws of probability	Semantically constrained Bayesian network
$P(d_i \mid m_j)P(m_j) + P(d_i \mid \neg m_j)P(\neg m_j) = P(d_i)$	$\boldsymbol{I}_{d_i m_j a=1}\boldsymbol{B}_{m_j} + \boldsymbol{I}_{\acute{d}_i m_j a=1}\boldsymbol{B}^c_{m_j} \leq \boldsymbol{B}_{d_i}$
$P(\neg d_i \mid m_j)P(m_j) + P(\neg d_i \mid \neg m_j)P(\neg m_j) = P(\neg d_i)$	$\boldsymbol{I}_{d_i m_j a=0}\boldsymbol{B}_{m_j} + \boldsymbol{I}_{\acute{d}_i m_j a=0}\boldsymbol{B}^c_{m_j} \leq \boldsymbol{B}^c_{d_i}$
$P(m_j \mid r_k)P(r_k) + P(m_j \mid \neg r_k)P(\neg r_k) = P(m_j)$	$\boldsymbol{I}_{m_j r_k a=1}\boldsymbol{B}_{r_k} + \boldsymbol{I}_{\acute{m}_j r_k a=1}\boldsymbol{B}^c_{r_k} \leq \boldsymbol{B}_{m_k}$
$P(\neg m_j \mid r_k)P(r_k) + P(\neg m_j \mid \neg r_k)P(\neg r_k) = P(\neg m_j)$	$\boldsymbol{I}_{m_j r_k a=0}\boldsymbol{B}_{r_k} + \boldsymbol{I}_{\acute{m}_j r_k a=0}\boldsymbol{B}^c_{r_k} \leq \boldsymbol{B}^c_{m_k}$

Table 3.

3.2.5. *Initial guess for influence factors*

The discussion on the laws of probability involving belief values and complementary belief values also holds for the influence factors. In the case of influence factors, the uncertainty amounts to only 40% i.e. less than the value associated with the degree of uncertainty for belief vectors. In the opinion of the authors, this is because a greater certainty can be attributed in this case as the occurrence of the metacause or rootcause nodes, respectively, is assumed. Whereas for the case of adjoint influence factors, the non-occurrence of the metacause or rootcause node, respectively, is assumed, hence the degree of uncertainty would be approximately 60%.

This yields the following modification in the laws of probability.

Laws of probability	Semantically constrained Bayesian network
$P(d_i \mid m_j) + P(\neg d_i \mid m_j) = 1$	$\boldsymbol{I}_{d_i m_j a=1} + \boldsymbol{I}_{d_i m_j a=0} \approx 0{\cdot}55 - 0{\cdot}6$
$P(d_i \mid \neg m_j) + P(\neg d_i \mid \neg m_j) = 1$	$\boldsymbol{I}'_{d_i m_j a=1} + \boldsymbol{I}'_{d_i m_j a=0} \approx 0{\cdot}35 - 0{\cdot}4$
$P(m_j \mid r_k) + P(\neg m_j \mid r_k) = 1$	$\boldsymbol{I}_{m_j r_k a=1} + \boldsymbol{I}_{m_j r_k a=0} \approx 0{\cdot}55 - 0{\cdot}6$
$P(m_j \mid \neg r_k) + P(\neg m_j \mid \neg r_k) = 1$	$\boldsymbol{I}'_{m_j r_k a=1} + \boldsymbol{I}'_{m_j r_k a=0} \approx 0{\cdot}35 - 0{\cdot}4$

Table 4.

4. The feasible set

In the light of the previous section it is evident that rigorous consistency checks need to be developed if this modified probabilistic approach is to work properly. Generating consistent and accurate probability values has always been a bottleneck in all such decision making approaches. In manufacturing processes such as sand casting, pressure die casting or injection moulding, the total number of probability values required would be of the order of thousands. In the particular case of pressure die casting process with 14 defect nodes, 18 metacause nodes and 43 rootcause nodes the total number of connections are around 250. (Assuming that one defect node is connected to at least 5 metacause nodes and one metacause node to at least 10 rootcause nodes.) The cost to be paid for the generalization of these probability values is four independent values get associated with each link in the network rather than the two. Also, two independent values are associated with each node rather than just one in the case of probabilistic analysis. As a result the program needs to check more than 1000 values which quantify the causal relationship. Certainly it is beyond doubt that it is unrealistic to expect production personnel to estimate all these values.

It is much easier to estimate a range for the likelihood ratios $(\boldsymbol{I}_{d_i m_j a=1}/\boldsymbol{I}'_{d_i m_j a=1})$ and $(\boldsymbol{I}_{d_i m_j a=0}/\boldsymbol{I}'_{d_i m_j a=0})$ respectively and an approximate guess for $\boldsymbol{I}_{d_i m_j a=1}$. The other values required to be guessed are $\boldsymbol{B}_{d_i}$ and $\boldsymbol{B}_{m_j}$, respectively to quantify a defect-metacause relationship. The program then generates, automatically, suitable values for $\boldsymbol{I}'_{d_i m_j a=1}$, $\boldsymbol{I}_{d_i m_j a=0}$, $\boldsymbol{I}'_{d_i m_j a=0}$, $\boldsymbol{B}^c_{d_i}$ and $\boldsymbol{B}^c_{m_j}$ such that all the values are globally consistent. Similar discussion holds for a metacause-rootcause relationship. This enumeration of consistent values is done based on the constraints developed earlier along with some more heuristics. Thus, a defect-metacause-rootcause relationship would be represented by only the following five values: $\boldsymbol{B}_{d_i}$, $\boldsymbol{I}_{d_i m_j a=1}$, $\boldsymbol{B}_{m_j}$, $\boldsymbol{I}_{m_j r_k a=1}$, $\boldsymbol{B}_{r_k}$ and all other values would be generated by the program consistent with these guesses. A particular assignment of consistent values to all these variables in a network represents one instance of the network. The feasible set defined in the following way stores all the possible instances of the network. Then a criterion is presented which selects only a couple of instances as the most plausible. The diagnosis is then done on the given example set using the selected instances of the network and the best instance is determined by judging the performance of the network on the example set. That particular instance is then saved and used for further diagnoses until any modification is deemed to be necessary.

For the sake of mathematical convenience, the feasible set is defined in the following way.

Definition:

The set of n-tuples $\langle \boldsymbol{B}_{d_i}, \boldsymbol{I}_{d_i m_j a=1}, \boldsymbol{B}_{m_j}, \boldsymbol{I}_{m_j r_k a=1}, \boldsymbol{B}_{r_k} \rangle, \forall i, j, k$ where $i = 1, \ldots, o$; $j = 1, \ldots, p$; $k = 1, \ldots, q$; $n = o + p + q + op + pq$, satisfying the following constraints is defined as a feasible set $\boldsymbol{FB}$.

Constraint set I:

$$1 \le \alpha_1 \le \boldsymbol{I}_{d_i m_j a=1} / \boldsymbol{I}'_{d_i m_j a=1} \le \beta_1 \le \infty \quad 1 \le \alpha_2 \le \boldsymbol{I}_{m_j r_k a=1} / \boldsymbol{I}'_{m_j r_k a=1} \le \beta_2 \le \infty$$

$$1 \le \alpha_3 \le \boldsymbol{I}_{d_i m_j a=0} / \boldsymbol{I}'_{d_i m_j a=0} \le \beta_3 \le \infty \quad 1 \le \alpha_4 \le \boldsymbol{I}_{m_j r_k a=0} / \boldsymbol{I}'_{m_j r_k a=0} \le \beta_4 \le \infty$$

$$\boldsymbol{I}_{d_i m_j a=1} + \boldsymbol{I}_{d_i m_j a=0} \approx 0{\cdot}55 - 0{\cdot}6 \quad \boldsymbol{I}_{m_j r_k a=1} + \boldsymbol{I}_{m_j r_k a=0} \approx 0{\cdot}55 - 0{\cdot}6$$

$$\boldsymbol{I}'_{d_i m_j a=1} + \boldsymbol{I}'_{d_i m_j a=0} \approx 0{\cdot}35 - 0{\cdot}4 \quad \boldsymbol{I}'_{m_j r_k a=1} + \boldsymbol{I}'_{m_j r_k a=0} \approx 0{\cdot}35 - 0{\cdot}4$$

$$\boldsymbol{B}_{d_i} + \boldsymbol{B}^c_{d_i} \approx 0{\cdot}45 - 0{\cdot}5 \quad \boldsymbol{B}_{m_j} + \boldsymbol{B}^c_{m_j} \approx 0{\cdot}45 - 0{\cdot}5$$

$$\boldsymbol{B}_{r_k} + \boldsymbol{B}^c_{r_k} \approx 0{\cdot}45 - 0{\cdot}5$$

Constraint set II:

$$0 \le \boldsymbol{I}'_{d_i m_j a=1} \le \boldsymbol{B}_{d_i} \le \boldsymbol{I}_{d_i m_j a=1} \le 1 \quad 0 \le \boldsymbol{I}'_{m_j r_k a=1} \le \boldsymbol{B}_{m_j} \le \boldsymbol{I}_{m_j r_k a=1} \le 1$$

$$0 \le \boldsymbol{I}'_{d_i m_j a=0} \le \boldsymbol{B}^c_{d_i} \le \boldsymbol{I}_{d_i m_j a=0} \le 1 \quad 0 \le \boldsymbol{I}'_{m_j r_k a=0} \le \boldsymbol{B}^c_{m_j} \le \boldsymbol{I}_{m_j r_k a=0} \le 1$$

$$\boldsymbol{B}_{d_i} \ge \boldsymbol{I}_{d_i m_j a=1} \boldsymbol{B}_{m_j} + \boldsymbol{I}'_{d_i m_j a=1} \boldsymbol{B}^c_{m_j} \quad \boldsymbol{B}_{m_j} \ge \boldsymbol{I}_{m_j r_k a=1} \boldsymbol{B}_{r_k} + \boldsymbol{I}'_{m_j r_k a=1} \boldsymbol{B}^c_{r_k}$$

$$\boldsymbol{B}^c_{d_i} \ge \boldsymbol{I}_{d_i m_j a=0} \boldsymbol{B}_{m_j} + \boldsymbol{I}'_{d_i m_j a=0} \boldsymbol{B}^c_{m_j} \quad \boldsymbol{B}^c_{m_j} \ge \boldsymbol{I}_{m_j r_k a=0} \boldsymbol{B}_{r_k} + \boldsymbol{I}'_{m_j r_k a=0} \boldsymbol{B}^c_{r_k}$$

where α_i and $\beta_1 \in \boldsymbol{R}$ are user defined.

The problem of generating the feasible set and finding the best instance of the network can be viewed as a general optimization problem. However, such general treatment does not appear promising as the total number of equality and inequality constraints total to around 3000. It is also not very easy to define a suitable objective function. Therefore, the following simple and useful modified backtrack algorithm is suggested to enumerate the consistent values to generate the feasible set.

A simple way to enumerate the consistent belief values and influence factors:

Step 1. From the initial guess of $\boldsymbol{B}_{d_i}$, $\boldsymbol{B}_{m_j}$, $\boldsymbol{B}_{r_k}$, $\boldsymbol{I}_{d_i m_j a=1}$ and $\boldsymbol{I}_{m_j r_k a=1}$ generate couple of prospective candidates close to the given initial guess (e.g. for the initial guess of 0·4, the program would generate values 0·3, 0·35, 0·4, 0·45, 0·5). The enumerator would attempt to assign the consistent values to the rest of the variables.

Step 2. For each value of the initial guess, use the constraint set I to generate couple of possible values (say 4 to 5) for each of the remaining variables.

Step 3. Delete those values for each variable which does not satisfy the constraint set II. This would generate the feasible set.

Step 4. For each set of values in the feasible set, test the performance of the diagnostic algorithm on the available example sets and select the best set of values as the final set.

5. The diagnostic algorithm

In the last section an enumerator is developed which will supply the consistent values enabling an execution of the diagnosis. The performance of the network should be checked on the given example set to decide which instance of the network gives the best results. The best instance of the network should then be stored and used for future diagnoses.

The diagnostic algorithm can be perceived as a generalization of the Bayesian analysis. The notions of likelihood ratio, and prior odds are generalized to incorporate all the enhancements as well as complexities discussed earlier in the paper. Again, the probability theory is modified in the following way.

Laws of probability	Semantically constrained Bayesian network
$O(H) = \frac{P(H)}{1 - P(H)}$	$\theta_j = \boldsymbol{B}_j / \boldsymbol{B}_j^c$
$L(e\|H) = \frac{P(e\|H)}{P(e\|\neg H)}$	$\boldsymbol{L}_{ij}^* = \frac{I_{ija}}{I_{i\bar{j}a}}$
Occurrence or non-occurrence of a hypothesis (H) $\propto P(H \mid e_1, e_2, e_3, \ldots, e_n)$	Occurrence or non-occurrence of a node $\propto \boldsymbol{A}_{out}(a_{in,j})$

Table 5.

However, the similarity to the Bayesian analysis ends here. As seen in the third point, the decision on the occurrence or non-occurrence of a particular metacause or rootcause node is not based on the generalized posterior odds. Instead, the notions of input activation and output activation of the node are introduced. The input activation to a metacause or rootcause mode is defined as the product of generalized likelihood ratios of all the preceding links. Figure 2 shows the way activation is computed in a successor node.

$\boldsymbol{A}_{in}$ is the input activation function which combines weighted activations from the earlier layers to decide on the activation of the node under consideration.

$$\boldsymbol{A}_{in}(a_i, \boldsymbol{I}_{ija=1}, \boldsymbol{I}_{i\bar{j}a=1}, \boldsymbol{I}_{ija=0}, \boldsymbol{I}_{i\bar{j}a=0}) = \prod_i \boldsymbol{L}_{ij}^*(a_i, \boldsymbol{I}_{ija=1}, \boldsymbol{I}_{i\bar{j}a=1}, \boldsymbol{I}_{ija=0}, \boldsymbol{I}_{i\bar{j}a=0}) \tag{3}$$

where $\boldsymbol{L}_{ij}^*$ is a generalized likelihood ratio for a link between nodes i and j. This is defined as follows (Fig. 3):

$$\boldsymbol{L}_{ij}^* = \begin{cases} \boldsymbol{L}_{ij}^c & \text{if } 0 \le a_i < 0{\cdot}1 \\ \boldsymbol{L}_{ij}^c + (1 - \boldsymbol{L}_{ij}^c)(a_i - 0{\cdot}1)/0{\cdot}3 & \text{if } 0{\cdot}1 \le a_i \le 0{\cdot}4 \\ 1{\cdot}0 & \text{if } 0{\cdot}4 < a_i < 0{\cdot}6 \\ 1 + (\boldsymbol{L}_{ij} - 1)(a_i - 0{\cdot}6)/0{\cdot}3 & \text{if } 0{\cdot}6 \le a_i < 0{\cdot}9 \\ \boldsymbol{L}_{ij} & \text{if } 0{\cdot}9 \le a_i \le 1{\cdot}0 \end{cases} \tag{4}$$

where $\boldsymbol{L}_{ij}^c = \frac{\boldsymbol{I}_{e_i h_j a=0}}{\boldsymbol{I}_{\bar{e}_i h_j a=0}}$ and $\boldsymbol{L}_{ij} = \frac{\boldsymbol{I}_{e_i h_j a=1}}{\boldsymbol{I}_{e_i h_j a=1}}$.

The output activation function converts the impulse $a_{in,j}$ into the activation level for the node under consideration. When the output activation of a node approaches unity, it is believed as to be occurring. Thus the occurrence and non-occurrence of a node gets scaled between 0 to 1.

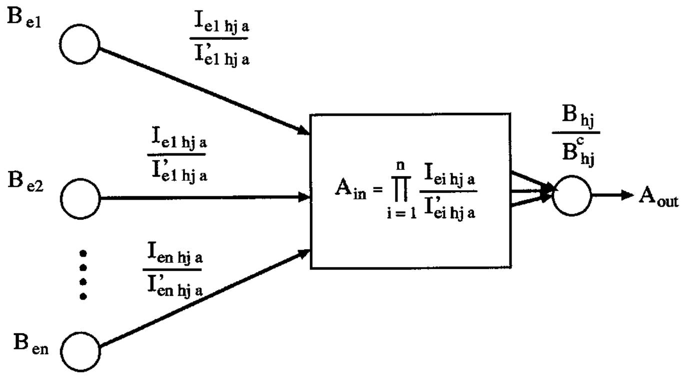

Figure 2. Activation in a successor node.

$$
\boldsymbol{A}_{out}(a_{in,j}) = \begin{cases} 1{\cdot}0 & \text{if } 5{\cdot}0 \le a_{in,j} < \infty \\ 0{\cdot}95 & \text{if } 4{\cdot}5 \le a_{in,j} < 5{\cdot}0 \\ 0{\cdot}90 & \text{if } 4{\cdot}0 \le a_{in,j} < 4{\cdot}5 \\ 0{\cdot}85 & \text{if } 3{\cdot}5 \le a_{in,j} < 4{\cdot}0 \\ 0{\cdot}80 & \text{if } 3{\cdot}0 \le a_{in,j} < 3{\cdot}5 \\ 0{\cdot}75 & \text{if } 2{\cdot}5 \le a_{in,j} < 3{\cdot}0 \\ 0{\cdot}70 & \text{if } 2{\cdot}0 \le a_{in,j} < 2{\cdot}5 \\ 0{\cdot}65 & \text{if } 1{\cdot}5 \le a_{in,j} < 2{\cdot}0 \\ 0{\cdot}60 & \text{if } 1{\cdot}0 < a_{in,j} < 1{\cdot}5 \\ 0{\cdot}50 & \text{if } a_{in,j} = 1{\cdot}0 \\ 0{\cdot}40 & \text{if } 0{\cdot}95 \le a_{in,j} < 1{\cdot}0 \\ 0{\cdot}35 & \text{if } 0{\cdot}85 \le a_{in,j} < 0{\cdot}95 \\ 0{\cdot}30 & \text{if } 0{\cdot}75 \le a_{in,j} < 0{\cdot}85 \\ 0{\cdot}20 & \text{if } 0{\cdot}65 < a_{in,j} < 0{\cdot}75 \\ 0{\cdot}10 & \text{if } 0{\cdot}55 \le a_{in,j} < 0{\cdot}65 \\ 0{\cdot}0 & \text{if } 0{\cdot}0 \le a_{in,j} < 0{\cdot}55 \end{cases} \tag{5}
$$

The posterior belief values are calculated only for user's information. For the jth node $\boldsymbol{PB}_j$, it is defined as follows:

$$
\boldsymbol{PB}_j = \boldsymbol{O}(a_{in,j}, \theta_j) \tag{6}
$$

where θ_j is the generalized posterior odds for the node j.

$$
\theta_j = \boldsymbol{B}_j / \boldsymbol{B}_j^c \tag{7}
$$

$$
\boldsymbol{O}(a_{in,j}, \theta) = (a_{in,j} \times \theta_j)/(1 + (a_{in,j} \times \theta_j) \tag{8}
$$

6. Validation

The validation of the semantically constrained Bayesian network is actually implied by the successful validation of ICADA (Ransing *et al.* 1995) in a realistic

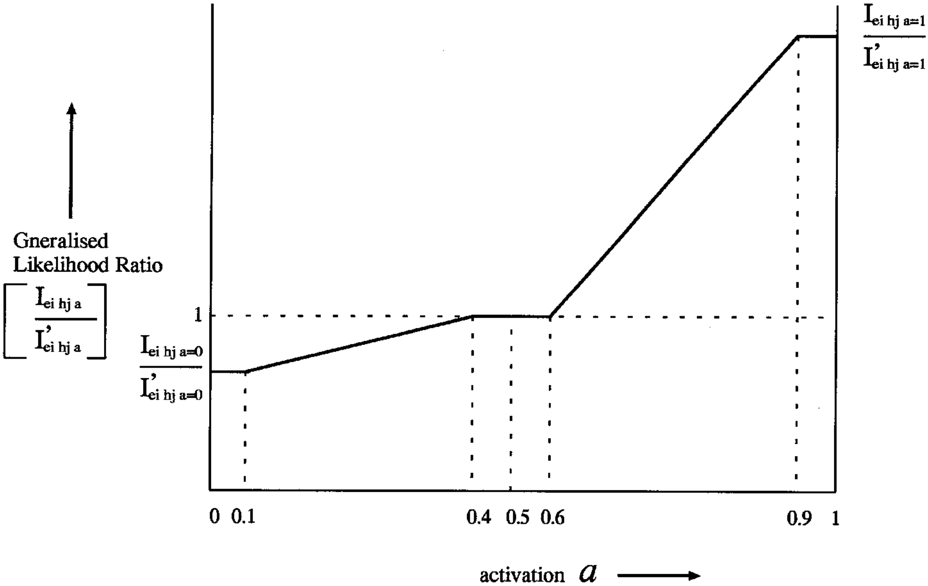

Figure 3. Generalized likelihood ratio variation.

environment. Because, it not only retains all the advantages of ICADA due to its innovative knowledge elicitation scheme but also offers following extra capabilities.

- Generation of consistent numerical values is automatic.
- Partial evidence is also considered during the analysis.
- More flexibility is available while calculating the posterior belief in the case of conflicting evidence, leading to a better control on the belief revision process.
- The constraints are more rigorous.

The best instance of the network was selected which gave the most satisfactory diagnoses on the given example set constituting 25 examples for a pressure die casting process manufacturing aluminium castings. This network, equipped with the most suitable belief values and influence factors, was then validated on another test data set comprising 20 diagnostic examples. One typical example has been illustrated below.

In all a total of 14 defect, 18 metacause and 43 rootcause nodes were identified. From the rejection data the input vector of activations was determined and is given as

$$a_{defect} = \left[0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1{\cdot}0\ 1{\cdot}0\ 0\ 0\ 0\ 0\right]$$

The activation of defect nodes *Build Up* and *Dimensional* is 1. The remaining defects do not occur i.e. *Porosity, Pin Holes, Blisters, Mismakes, Stains, Cold Shuts, Drags, Cracks, Flash Overlap, Die Damage, Bending* and *Broken Core*. For this input vector the activation in metacause nodes was:

$$a_{metacause} = \left[0\ 0{\cdot}75\ 0\ 0{\cdot}6\ 0{\cdot}75\ 0{\cdot}6\ 0\ 0{\cdot}6\ 0\ 0{\cdot}6\ 0\ 0\ 0{\cdot}65\ 0{\cdot}65\ 1{\cdot}0\ 0\ 0\ 0\right]$$

The nodes with activation 0·75 and above are *cooling rate of casting–inadequate* (0·75), *Lubrication–Inadequate* (0·75), *Overheating of the die* (1·0). Some of the nodes show some positive indication whereas others are with zero activation. Activations in rootcause nodes are then determined. Rootcauses with activation close to unity (values in the round brackets) were as follows:

- Cycle time (1·0) with posterior belief—95%
- Quantity of die lubricant – less (0·95)—98%
- Metal too hot (0·8) with posterior belief—85%
- Mixing ratio of die lubricant (0·7) with posterior belief—90%
- Temperature and flow rate of the cooling system (0·7) with posterior belief—84%
- Number of the cavities (0·7) with posterior belief—89%
- Kind and number of cooling systems (0·65) with posterior belief—78%

The production personnel agreed to scrutinize the following three causes first:

- Whether the cycle time was correct.
- Temperature of the molten metal.
- Temperature of the die, whether it cooled adequately or not.

Remarks:

(1) It is to be noted that the activation decides the occurrence of a particular node and not the posterior belief as in the case of ICADA.
(2) As a result of more control on the belief revision, the diagnosis can closely match reality.
(3) The semantically constrained Bayesian network has been validated on 20 representative data sets with different dies and time frames.

The sample calculation of the activation in the metacause node *Overheating of the die* is given below.

Defect nodes	a_i	$\boldsymbol{I}_{ija=1}/\boldsymbol{I}_{\acute{i}ja=1} = \boldsymbol{L}_{ij}$	$\boldsymbol{I}_{ija=0}/\boldsymbol{I}_{\acute{i}ja=0} = \boldsymbol{L}^{c}_{ij}$	$\boldsymbol{L}^{*}_{ij}$
Porosity	0	0·44/0·44 = 1	0·16/0·16 = 1	1
Pin holes	0	0·34/0·10 = 3·4	0·26/0·31 = 0·83	0·85
Blisters	0	0·14/0·14 = 1	0·36/0·36 = 1	1
Mismakes	0	0·32/0·32 = 1	0·18/0·18 = 1	1
Stains	0	0·11/0·11 = 1	0·39/0·39 = 1	1
Cold shuts	0	0·2/0·2 = 1	0·3/0·3 = 1	1
Drags	0	0·26/0·26 = 1	0·24/0·24 = 1	1
Cracks	0	0·32/0·32 = 1	0·18/0·18 = 1	1
Build up	1	0·30/0·07 = 4·29	0·24/0·30 = 0·8	4·29
Dimensional	1	0·42/0·05 = 8·40	0·18/0·26 = 0·69	8·4
Flash overlap	0	0·29/0·29 = 1	0·21/0·21 = 1	1
Die damage	0	0·28/0·04 = 7	0·28/0·37 = 0·76	0·75
Bending	0	0·26/0·26 = 1	0·24/0·24 = 1	1
Broken core	0	0·20/0·20 = 1	0·30/0·30 = 1	1

Table 6.

The activation and the posterior belief are calculated as follows:

$$\begin{aligned} a_{in,j} &= \prod_i \boldsymbol{L}^*_{ij}(a_i, \boldsymbol{I}_{ija=1}, \boldsymbol{I}'_{ija=1}, \boldsymbol{I}_{ija=0}, \boldsymbol{I}'_{ija=0} \\ &= 0{\cdot}85 \times 4{\cdot}29 \times 8{\cdot}4 \times 0{\cdot}75 \times 1 \\ &= 22{\cdot}97 \end{aligned}$$

$$\begin{aligned} a_j &= \boldsymbol{A}_{out}(a_{in,j}) \\ &= 1{\cdot}0 \end{aligned}$$

$$\begin{aligned} \theta_j &= \boldsymbol{B}_j / \boldsymbol{B}^c_j \\ &= 0{\cdot}29/0{\cdot}21 \\ &= 1{\cdot}38 \end{aligned}$$

$$\begin{aligned} \boldsymbol{PB}_j &= \boldsymbol{O}(a_{in,j}, \theta_j) \\ &= (a_{in,j} \times \theta_j)/(1 + (a_{in,j} \times \theta_j) \\ &= (22{\cdot}97 \times 1{\cdot}38)/(1 + (22{\cdot}97 \times 1{\cdot}38)) \\ &= 0{\cdot}97 \end{aligned}$$

7. Conclusions

The semantically constrained Bayesian network is proposed as a diagnostic tool. The overall objective in developing this network can be categorized as follows:

Represent causal relationship in the form of a directed network such that it is easy to quantify the links with local, conceptually meaningful parameters that turn the network as a whole into a globally consistent knowledge base.

The notion of Bayesian analysis has been utilized while quantifying the causal relationship as well as the belief revision process (i.e. estimating activation in the respective nodes.) However, instead of assessing the probability of meta/root cause being true, the level of confidence that can be imparted in a meta/root cause is assessed and estimated. As a result, the complete knowledge regarding causal relationships – a must in the case of probabilistic models – is not necessary.

Acknowledgments

The partial financial support through an European Grant COST 504 is gratefully acknowledged. The authors would like to thank the management of Kaye Presteigne Ltd., UK, for allowing discussions with their personnel and providing realistic data for validation.

References

Creese, R. C., 1988, Introduction to expert systems to foundry applications. *AFS Transactions*, **96**, 443–446.

Davis, R., Buchanan, B. G., and Shortliffe, E. H., 1977, Production rules as representation for a knowledge-based consultation program. *Artificial Intelligence*, **8**, 15–45.

Fujikawa, S., and Ishi, K., 1995, Diagnostic expert systems for defects in forged parts. *Journal of Intelligent Manufacturing*, **6**, 163–173.

Hentenruyck, P. V., 1989, *Constraint Satisfaction in Logic Programming* (Cambridge, MA: MIT Press).

Juran, J. M., 1964, *Managerial Breakthrough* (New York: McGraw-Hill).

Kuo, T., and Mital, A., 1993, Quality control expert systems: a review of pertinent literature. *Journal of Intelligent Manufacturing*, **4**, 245–257.

Pandelidis, I. O., and Kao, J. F., 1990, DETECTOR: A knowledge-based system for injection molding diagnostics. *Journal of Intelligent Manufacturing*, **1**, 49–58.

Patterson, D. W., 1990, *Introduction to Artificial Intelligence and Expert Systems* (Englewood Cliffs, NJ: Prentice Hall).

Pedrycz, W., 1993, Principles and methodology of fuzzy sets. *Journal of Intelligent Manufacturing*, **4**, 323–340.

Piwonka, T. S., 1989, Current and potential use of process modeling for foundry process control. *AFS Transactions*, **97**, 465–472.

Ransing, R. S., Srinivasan, M. N., and Lewis, R. W., 1995, ICADA: Intelligent computer aided defect analysis for castings. *Journal of Intelligent Manufacturing*, **6** (1), 29–40.

Shafer, G. A., 1979, *Mathematical Theory of Evidence* (Princeton: Princeton University Press).

Shortliffe, E. H., 1976, *Computer Based Medical Consultations: MYCIN* (New York: Elsevier).

Ward, T. L., Ralston, P. A. S., Karwowski, W., and Hall, W. D., 1991, ITONUS: expert system for machining on a lathe. *Journal of Intelligent Manufacturing*, **2**, 353–363.

Zadeh, L. A., 1965, Fuzzy sets. *Information and Control*, **8**, 338–353.

Zadeh, L. A., 1978, Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, **1**, 3–28.