SRC TR 87-198

Priority-First Minimum Cover Approach to Diagnostics with Application to Injection Molding

by

I. Pandelidis and Z. Lin

Priority-First Minimum Cover Approach
to Diagnostics with application to injection molding

Abstract

A new methodology is being proposed for application to diagnostics in manufacturing problems. The approach adopts a minimum cover approach where we propose that the requirements for priority and consistency as will be later defined are at least as important as minimum-cover. Thus a priority-first minimum set covering model is proposed to deal with some diagnostic problems, particularly related to manufacturing defect diagnostics. This approach is applied to implement a diagnostic system for manufacturing problems in injection molding.

I. Introduction

Expert systems behave satisfactorily in medical diagnosis. In recent years,much work has been done to explore the potential of expert diagnostic systems in computer aid manufacturing. The basic idea is to integrate the valuable human expertise into a computer program to be utilized by non-expert workers. Hence the quality of products, and efficiency of manufacturing could be improved without the presence of a human expert.

The concept of Artificial Intelligence has existed for several decades. But only until recent years has it become a practical for industrial applications. Many manufacturing problems are solved by relying primarily on expertise rather than science. Expertise usually is not easy to be abstracted or formalized so as to be "understood" or "manipulated" by today's computers. One key difficulty lies in the fact that we may not always be at a position to develop theory . Another difficuly arises when the experts themselves may not be able to communicate their knowledge to others.

Another problem we will be facing is the uncertainty of expertise,in the sense that rules may apply to a significant portion of cases at hand but associated with a level of confidence. The methodology to deal with this issue has been explored and is still under study. One obviously desirable way is to improve the quality of the rules of experience. But due to the reasons mentioned above, this would always turn out to be somewhat impractical. So it leaves open the question of how to improve the performance of an expert diagnostic system by taking into account the uncertainty of specific rules.

Our project of injection molding diagnostic system addresses the above issues. The diagnostic program in this system will take human description of plastic product defects as input, output the diagnosis and suggest a remedy to those defects. Other functions such as explanation and discussion are also implemented.

Section II presents a functional overview of the diagnostic system; section III discusses priority-first minimum cover theory; section IV describes some methodology used in implementing the system, and demonstrates a couple of test cases.The conclusions are given in section V.

II.1 System Structure

The system is organized in a way that it is centered by a knowledge base. The knowledge base consists of stored experiential data, theoretical rules of diagnosis, logic reference rules, technical data of general materials and injection machines, as well as rules monitoring the activity of the system itself. Around the knowledge base, three interactive functional components are built to provide facilities for diagnosis, explanation and discussion. On-line help is also implemented. Figure 1. illustrates the overall system graphically.

Figure 1.

II.2 Functional Components

A. Knowledge base:

There are five functionally independent parts of this knowledge base. They are:

1). Trouble shooting guide: query will first be directed to this table for quick solution.

2). Reference rules: this part contains general rules about the properties and relationships among properties of plastic materials. These rules will be invoked as soon as the trouble shooting guide fails to identify the causes of certain defects.

3). Relation of plastic variables and machine variables: Remedy to the defects will be given after the diagnosis is achieved. Suggested remedy will be in terms of the adjustable variables or machine variables, while causes would in most cases be in terms of plastic variables. **have not explained what the problem is in injection molding brief description required**

4). Materials and Machine data: technical data of common materials and machine, such as process temperature, injection pressure, etc. It will also contain experiential data of some local operating environment. This data will contribute towards making a precise judgement.

5). Dictionary: a simple dictionary is stored for use of logic reference and language comprehension.

B). Diagnostic program

A detailed description of this program will be given in the next section.

C). Explanation program

Inputs to this module are a list of defects and associated causes. The program will trace and review all the rules invoked during the diagnostic process giving the appropriate explanation of how the defects are related to the causes.

D). Discussion program

This program invokes the diagnosis program again after knowing that some of the causes given by previous diagnosis are denied or confirmed, thus yielding another diagnosis based on the new information. As should be expected, more accurate conclusions should be drawn when more conditions are known.

II.3 Operations

The user interface is formed by the diagnostic, explanation and discussion programs described above, plus HELP facility. So far the interface is command-driven, but it can be easily converted to be menu-driven, depending on the type of users it is serving. Basic commands and command format are as following:

command	**operation**
reset:	All variables in the system are reset to default values.
diagnose:	Invoke the diagnostic program. This program will first ask for defects description, then it looks into the trouble shooting guide in the knowledge base for a solution. If it succeeds, which means a number of possible causes for the given defect(s) are found, it will proceed on to the CONFIRMING procedure, otherwise it will look for proper reference rules in the knowledge base and try to deduce the causes for the defect(s), then enters the CONFIRMING procedure. The CONFIRMING procedure takes those possible causes as input, processes these data and generates a priority-first minimum cover for the defect(s), which we consider as the most possible causes to

the given defect(s). After this the program proceeds on to the REMEDY procedure, in that the suggested remedy to the most possible causes are found and returned to the user.

explain: Invoke the explanation program. This program will ask for defect and causes that need an explanation, then backtrace the diagnostic process and return, in a legible form, the rules used in that process in the order they were invoked, as an explanation for how the causes are related to the defects.

discuss: Invoke the discussion program. This program allows the user to argue with the system the results it achieved, such as confirming some particular causes or denying the other. Diagnostic program will be invoked again using the new information, and new diagnosis will be returned.

help: Some kind of on-line help for system overview, command syntax and description, and guide to using the system.

III. PRIORITY-FIRST MINIMUM SET COVERING MODEL

Minimum set covering model was studied and proposed by some other researchers [1] as a general model for medical diagnostic problem solving, particularly for problems where multiple disorders exist. Unfortunately we can not adopt directly this approach in our system, due to reasons that first this model assumes the independence of diseases(evidences), which can not be achieved in our problem of injection molding, and second,this model neglects the importance of the fact that diseases(evidences) usually occur with different probabilities for the given disorder(hypothesis), but this we think should be considered even more important than the minimalization of the number of the evidences, although we have no doubt about the basic assumption, known as the Principle of Parsimony of Ockham's Razor, "The simplest explanation is the preferable one".

Having all this in mind we proceeded onto an alternative direction, which leads to the development of the priority first minimum set covering model. It is the objective of this model to deal with the problem of ambiguity which widely exists in both experiential rules and evidence that explains multiple disorders, especially in the diagnostic problems in injection molding. This model was implemented in our diagnostic system with very encouraging results. Details of this model will be described in following.

III.1 Priority-first set covering model

The diagnostic problem is generally understood to be that by which a certain description of disorders(defects) is given and a cause (disease) are required to be found. In addition a remedy (prescription) is also required to deal with the underlying cause. Explanation should be sufficient, in the sense that each disorder must have reasons behind it, and if possible, necessary **define what is meant by necessary **. The latter is a difficult part because it is heavily problem dependent and no accepted standard is presently available to judge necessity as such.In general the human diagnostician would prefer the simple explanation as the necessary one if no evidence to support that a more sophisticated one is better. And this assumption will be valid in our diagnostic process.

Yet there are at least two more important issues we have to take into account in diagnosis, particularly in the problem we are dealing with in injection molding. One is the weight, or importance, of each reason that explains the disorder. The other is the consistency among the reasons found to be supporting the disorders. Consistency here should be viewed in a syntactic as well as in a semantic way. Given a formalized way of representing all this knowledge, modules should be developed that will check for syntactic as well as possibly semantic consistency in a method independent of the particular process at hand.

Having considered all these issues, we are now ready to define the priority-first minimum set covering model, which reflects our concerns about priority, consistency, and simplicity.

DEF1. Given a set D of defects, a cover to D is a set C of causes such that for any defect d in D, there exists c in C, c causes d and any c_i, c_j in C, c_i and c_j are **consistent.**

Example 1.

Given a set D of defects $[d_1, d_2, d_3]$, suppose we know the following relation: (assume c_1, c_2, c_3, c_4 consistent with each other)

$$cause - of(d_1) = c_1, c_3;$$
$$cause - of(d_2) = c_1, c_2, c_4;$$
$$cause - of(d_3) = c_4.$$

The above may be represented graphically as :

d_1 d_2 d_3

c_1 c_2 c_3 c_4

A cover C to D is either $[c_1, c_4]$, or $[c_3, c_4]$, or $[c_1, c_2, c_4]$, or$[c_1, c_3, c_4]$, or $[c_2, c_3, c_4]$, or $[c_1, c_2, c_3, c_4]$.

DEF2.A cover C to set D of defects is **minimum** iff for any cover C' to D, $|C| <= |C'|$,where $|A|$ denotes the cardinality of set A.

Example 2.

In the above example, the minimum cover is either $[c_1, c_4]$, or $[c_3, c_4]$.

DEF3. The **weight** of a cause c_i to defect d_j is an integer w_{ij} .

The weight is in fact denoting the likehood of a cause to the defect that it relates to. Note that we do not require it is the probability of a cause, which has strict mathematical definitions that may not always apply to real applications.

Example 3.

In example 1, let:

$$weight(d_1, c_1) = 5;$$
$$weight(d_1, c_3) = 3;$$
$$weight(d_2, c_1) = 1;$$
$$weight(d_2, c_2) = 3;$$
$$weight(d_2, c_4) = 4;$$
$$weight(d_3, c_4) = 10;$$

DEF4. The **weight** of cover C to set D of defects is a function $W_c : [w_{ij} :$ where w_{ij} is weight of $c_i \in C$ to $d_j \in D.] -> \mathrm{R}$

A reasonable assumption to make is that Wc should be incremental. By that we mean that Wc does not decrease as the weight of each cause in the cover increases. Reference[2] supports this line of reasoning.

Example 4.

In order to demonstrate the concept let Wc be the accumulation of all the weights of c_i to d_j where $c_i \in C$ and $d_j \in D$. We have:

$$Wc([w_{11}, w_{12}, w_{42}, w_{43}]) = 20$$
$$Wc([w_{31}, w_{42}, w_{43}]) = 17$$
$$Wc([w_{11}, w_{12}, w_{22}, w_{42}, w_{43}]) = 23$$
$$Wc([w_{22}, w_{31}, w_{42}, w_{43}]) = 20$$
$$Wc([w_{11}, w_{12}, w_{31}, w_{42}, w_{43}]) = 23$$
$$Wc([w_{11}, w_{12}, w_{22}, w_{31}, w_{42}, w_{43}]) = 26$$

DEF5. A **priority-first minimum cover** to set D of defects is a cover C to D such that $W_c/|C|$ is maximum, where W_c is the weight of C.

Note that priority-first minimum cover is NOT a minimum cover! It is a cover with the average weight of its elements being maximum, as compared with all the other covers.

Example 5.

In the above example, the priority-first minimum cover to set D is $[c_1, c_4]$. Intuition would support the conclusion that this cover should be the best explanation to the defects $[d_1, d_2, d_3]$, with respect to its sufficiency and necessity.

III.2 Algorithm for finding priority-first minimum cover

Problem: Given a set D of defects, suppose the causes to D are known and stored in list "Diseases" in descendant order of the *weight* of each cause, which is calculated based on a particular method which would not concern us at the moment, find the priority-first minimum cover to D.**needs more clear description**

Method: We will prove later on that the priority-first minimum cover must include the first n elements in the ordered list of causes. So we will first find a *minimal* cover, which is approximately the priority first minimum cover to D, by searching along the causes list.

Input:

Defects: List of defects, correspondent to set D;

Diseases: Ordered list of causes to D;
Old: Cover found so far;

Output:

New: Final "*minimal*" cover.

Algorithm:

```
    confirm(Defects,Diseases,[],New).
confirm([],,, []) : - !. /* cover found */
confirm(D,[],,[]) : - /* see explanation [1] */
        write('Consistent cover not exists').
confirm(Defects,[Cs|Diseases],Old,[Cs|New]) : -
        consistent(Cs,Old), /* if Cs consistent with current result */
        crossout(Cs,Defects,Remain), /* see explanation [2] */
        !,
        confirm(Remain,Diseases,[Cs|Old],New).
confirm(Defects,[|Diseases],Old,New): -/* see explanation [3] */
        confirm(Defects,Diseases,Old,New).
```

Explanation:

1. Cover to arbitrary defects D may not always exist.
2. Defect(s) corresponding to Cs will be deleted in list Defects, the remaining list will be returned in Remain.
3. Inconsistent causes to the already found cover will be ignored because it has lower priority. This is a only a simplified scheme here as an example.

Now we prove the claim: "Given set D of defects, let C be the ordered list of causes to D, then there is a priority-first minimum cover that includes the first k elements in C for some $k > 0$, provided D has priority-first minimum cover."

Proof:

Let C be the ordered list of causes to set of defects D

$$C = [c_1, c_2, ..., c_n]$$

We have $w_i > w_j$ when $i > j$, where w_i is the weight of c_i

Proof by contradiction.

Suppose there is no priority-first minimum cover to D that includes the first k elements in list C for any $k > 0$.

Let

$$P_c = [c_{m1}, c_{m2}, ..., c_{mn}]$$

$$W_c = w_{m1} + w_{m2} + ... + w_{mn}$$

where P_c is a priority-first minimum cover to D, with *weihgt* being W_c, w_i is the correspondent weight of c_i and $w_i >= w_j$ when $i > j$, $mn > ... > m2 > m1 > k > 0$.

We have $|P_c| = n$

$W_c/|P_c| = (w_{m1} + w_{m2} + ... + w_{mn})/n$

Since $m1 > k >= 1$, c_1 is not in P_c

Let P'_c be union of P_c and $[c_1]$, W'_c be weight of P'_c and $W'_c = c_1 + W_c$. Note that P'_c is also a cover to D, $W'_c/|P'_c| = (w_1 + W_c)/(n+1) >= W_c/n$, since $w_1 >= w_{m1} >= w_{m2} >= ... >= w_{mn}$.

case 1. $W'_c/|P'_c| = W_c/n$, then the cover $[c_1, c_{m1} ... c_{mn}]$ is a priority-first minimum cover to D and it includes the first element of C. Contradict to hypothesis (k=1).

case 2. $W'_c/|P'_c| > W_c/n$, then P_c is not a priority-first minimum cover. Contradict to that P_c is a priority-first minimum cover.

End of proof.

The proof here suggests that if we are to find the priority-first minimum cover we could start by scanning from the head of the ordered list of causes, adding elements to the result list until a cover is found. This cover is not yet the desired cover but very close to.

The next step is to eliminated those elements in this cover that do not contribute to a maximum average weight, and add in elements which are not in the present cover but that would help increase the average weight. For clarity we omit such algorithm here.(However, we would suggest using the "confirm" procedure as an anologous algorithm for finding the PFMC, though it not a PFMC in the strict sense. We can show statistically there is little chance that the second step would significantly increases the average weight of the cover in most cases.)

IV Implementation of System

The priority-first minimum cover(called PFMC so after) model presented in the above section is basically a decision making model in diagnosis, as is necessary to help determine the most possible explanation among dozens of possible ones. Obviously this is not the only issue that has to be considered to build a diagnostic system.

The goal we were aiming at in our project is a system that will eliminate the user involvement in the diagnostic process to the most neccesary level and that would potientially to be adopted as part of the atuomative control system of injecton molding. Several important features have been addressed in implementing such a system, besides the PFMC model. And we intend to discuss them in the following.

IV.1 Feed back mechanism

Any "intelligent" system possesses certain kinds of feed back mechanism. Because it is so important a way to guarantee a correct output. Examples lie from the tiny electronic circuit to the sophysicate human nervours system. There is no reason why we did not copy such a model from nature.

As matter of fact, most of today's diagostic systems do have the hypothesis-test cycle, in that hypothesis is established by the system and tests are planed and sent to the user. Test results are "feed back" to the system as used to confirm or deny the hypothesis. We call it internal feed back mechanism because it happens during the diagnostic process.

But, as we mentioned before, we want a system that can work as independently as possible, in the sense that it requires least human involvement during its execution. So it is intentionally designed that the number of questions by the system to the user is kept to the smallest. However, this does not mean we abondan the feed back technique. It is just so important that we can not affort without it.

We see the solution in a alternative structure of feed back scheme, which we called external feed back mechanism. To understand this let us take a closer look to the system in Figure 2.

Figure 2.

Input data(defect description, material, machine description, etc) are received by PC(possible causes) generator, which detects all the possible causes to the given defects, and assigns to each cause a number(weight) denoting its contribution to the explanation of the defects. This number is calculated with regards to:

1) prior probability of the cause to every defect it related to;

2) Severity of defects;

3) Processing enviroment(machine setting, mold, machine condiction);

4) Difference between machine setting and experiential data for similar process.

The result of the PC generator, which is a list of causes sorted by their weights, is passed to the PFMC module. The PFMC module is a procedure based on the PFMC model. In this procedure, the PFMC to the given defects is selected, and the output is the explanation to the defects, which is actually the causes in the PFMC, and the possible causes, which are the causes detected by PC generator but not included in the PFMC.

The explanation to the defects is taken over by the RD(remedy) module, in that the remedy to the causes is planed and returned to the user.

Problem here is that the output of PC generator can not always be guaranteed correct, especially the weights of the causes. And these factors would effectively affect the solution of the whole diagnosis. Anticipating that this thing might happen, we must be prepared for it. It comes out to be, as suggested by the graph, a feed back ring which closes the PFMC module. In fact, this is the Discussion program. It takes the output of PFMC, and userjudgements as input, feed back this "signals" to the input of PFMC so as to "modify" the previous diagnosis. What actually happened inside the Discussion program is that the weights of the cause are adjusted according to the judgement given by the user.

It is a plausible scheme here to have such feed back structure because it provides the user with a flexible way to take part in the diagnosis. Note that the connection of Discussion program with the PFMC module is dynamic, which means the program will be invoked only on requests by the user. The latter would no longer be bothered by a screenful of questions he may not like, most of them was not funny, if not dump. Moreover, this is a structure that can easily incorporate the learning capability, because not only the weights of causes can be adjusted inside the Discussion program, but also the RULES or VARIABLES in the knowledge base: system could learn from its discussion with a expert user. But unfortunately this has not yet been implemented.

IV.2 Data of material properties

Data of plastic property, and their general processing condictions are stored in the knowledge base. They form a fairly independent division of the knowledge base, and are basically prepared for the use of the diagnostic program. Queries like "what value dose property P of material M has ?", and "what would influence property P of material M ?" will be responed by looking up the database.

Serious problem risen here is the redundancy of data. Resin of different grades usually has good amount of common properties. Obviously it is not the practice to store every single one of them. Instead, relations between properties should be made good use of. For example, the "inheritable" relation denotes what properties a subgrade can inherite from its parent. In our system, this relation was formalized into a hierarchical model. When query about a property arises, and that property is not stored as a fact in the database, it will be directed to the parent of the present subgrade, if it exists.

Similar approach were used to deal with the "influence" relation, which states for material M, property X will be influenced by property Y with proportion P, provided X is within range R. This is used for detecting problems with certain property. For example, the orientation of material 1001 is influenced by the melt tempreture and packing presure during injection, and orientation in turn influences the tensile strength of the product, so when we notice that tensile strength of the part is not adequate, we should be able to reason that melt tempreture and/or packing presure may be wrong, by following the "influence" rules.

Finally, we must point out that it is not our objective in the project to incorporate a complete database into the present diagnostic system. Works having been done so far are focused on developing the reference mechanism without regards of the physical, or even conceptual storage of the data. In other words, the material database here is in fact more of a user-oriented interface than a real database, which in our opinion needs not to be particularly designed for the facility of a diagnostic system. Actually this work is being done currently in our another project of ************.

bf IV.3 Test samples

We have already described the PFMC model and several implemetation details of the diagnostic system and now we believe is the time to show how the system works by some real test samples. The system is, or is supposed to be, characterized with:

1. Multiple defects handling;
2. Specific but arguable results;
3. Less user involvment;
4. Capable of self explanation;
5. Clear remedy suggestion.

We would see these features through the following examples.

Sample 1.

The first test is with a sample with obvious evidnece of bubbles and dimensional variation, while the operating envioremnent known to the computer is nothing abnormal(kind of the toughest test!). Results as following:

********* Diagnostic report **********

Defects description:

Obvious evidence of bubbles
Obvious evidence of dimensional-var

The operating enviorement as I know is:

The mold has no special evidence of abnormal.
The material has no special evidence of abnormal.
The machine has no special evidence of abnormal.

My diagnosis is:

low plastic-feed with relative weight 21 which causes defects:
bubbles of 10.5
dimensional-var of 10.5

Other possible causes may be:

uneven mold-temp with relative weight 13.5
uneven melt-temp with relative weight 13.5
irregular part-design with relative weight 13
high melt-temp with relative weight 10.5
low hold-time with relative weight 10.5
low fill-rate with relative weight 10.5
low plastic-cap with relative weight 7
incorrect gate-dim with relative weight 7

variation contamination with relative weight 7
wet contamination with relative weight 7
dirt contamination with relative weight 7
incorrect runner with relative weight 6
irregular particle with relative weight 6
low machine-cap with relative weight 6
incorrect gate-loc with relative weight 6
general distortion with relative weight 6
large runner with relative weight 5
small runner with relative weight 5
incorrect functioning with relative weight 5

The following action can be taken as to eliminate the causes I detected:

You can increase shot-size-or-injection-pre to increase plastic-feed.
But, plastic-feed may be too high.

Diagnosis finished and reuslts are stored in file "result". You can review this file by typing view(result).

If you do not agree with this result, you can discuss it with me by typing "discussion" at the command prompt.

If you want me to explain how this result is achieved, please type "explanation" at the command prompt.

*********** End of report ************

Comments:

This diagnostic record tells that the most possible cause to the defects of serious bubbles and dimensional variation is insufficient plastic feed. Results are based on the fact that machine settings, material and machine condictions have no evidence of abnormal, and machine and material being used are standard.(The latter information was stored in file "clinic" and read by the diagnostic program at its initialization phrase).This is a reasonable result as we can image, since bubbles and dimensional variation strongly suggest insufficient plastic feed as their cause seperately, though insufficient plastic feed is not the best explanation to the present of either bubbles or dimensional variation in the current condictions. Multiple defects help detect the REAL cause(s), which are not the simple accumulation of causes to every single defect. Also, all other possible causes are listed in the order of their weight to the explanation of the defects, in accordance to conventional approach. Suggested remedy is to increase shot size or increase injection pressure, which is quite to the point and easy to follow. Side effect is given as caution to applying the remedy.

Sample 2.

In the above sample, suppose we eventually find out that insufficient plastic feed is not the case(false diagnosis), we will then invoke the discussion program to let the system know the result is denied. The new diagnosis based on this information is given as following.

********* Diagnostic report **********

.

.

.

My diagnosis is:

uneven melt-temp with relative weight 13.5 which causes defects:
bubbles of 13.5
uneven mold-temp with relative weight 13.5 which causes defects:
dimensional-var 13.5

.

.

.

The following action can be taken as to eliminate the causes I detected:

You can maintain-gate-loc to maintain even melt tempreture.
You can maintain even-mold-temp to maintain even mold tempreture.

*********** End of report ************

Comments:

New diagnosis well explains the defects of bubbles and dimensiona variation, provided knowing insufficient plastic feed is not the cause. The last statement in report is a little bit awkard. But it should make sense to a operator since he should know mold tempreture is a directly adjustable varible.

Notice that the system assumes nothing has changed in the operating envriorement since the last diagnosis. If this is not the case, reset the system is necessary before any other transaction.

Summary

It was discussed in this paper a new diagnostic model which is constructed on the study of the previous works on relevent subjects and the current project of a diagnostic expert system for injection molding. It is intuitively a general model for most of the diagnostic problem solving in industry because of its capability of giving specific solution to multiple disorders problem with shadow experiential knowledge. The model is justifiable by our general understanding of human diagnostic process and partially justifiable by mathematical theory. We see its potential as major step towards the utilizing of an expert system in automative control in production.

However, this new model still evokes future tests in practice and theory. Several problems in implemetation were very challanging and they have not yet been satisfactorily solved. More specifically, they are the problem of ranking the weight of each cause and weight of a cover, the hypoheses-test scheme(if we insist it is necessary), and the internal consistency problem which will arise as soon as the learning ability is incorporated. We intend to continue our research on these issues.

Reference:

1. "Diagnostic expert system based on a set covering model". By J. A. Reggia, D. S. Nau and P. Y. Wang. Development of Expert System, 1984

2. "Subjective Inference with Multiple Evidence". By R. R. Kadesch. Artificial Intelligence 28(1986)333-341.

3. "Probabilistic Logic". By N. J. Nilsson. Artificial Intelligence 28(1986) 71-87.

4. "Molding Problems and Solutions". *INJECTION MOLDING handbook*

5. "Expert System in Injection Molding" *CAE*

6. "ABS *******"

7. "Nylon*******"

8. "Mycin******"

9. "PETS" *GE*