RELATIONSHIPS BETWEEN DEDUCTIVE AND ABDUCTIVE INFERENCE IN KNOWLEDGE-BASED DIAGNOSTIC PROBLEM SOLVING[1]

Dana S. Nau and James A. Reggia

Computer Science Dept.
University of Maryland
College Park, MD 20742

Abstract

Most knowledge-based computer systems are based on a production rule format which proves to be very difficult to apply to many diagnostic problems. In this paper, we argue that this is because diagnostic problem solving is a problem in abductive inference rather than deductive inference. The usual rule format does not provide sufficiently rich knowledge representation capabilities to allow abductive problems to be translated into deductive ones. We also show how abductive problems can be translated into deductive problems by using a rather different rule format than is usually used in knowledge-based computer systems, and discuss an algorithm for diagnostic problem solving based on this rule format. This algorithm has been successfully used to create expert computer systems in several different problem domains.

1. Introduction

The vast majority of expert systems for diagnostic problem solving which have been developed by artificial intelligence researchers are based on the use of production rules to do deductive inference [3]. Even systems such as Prospector [1] which use semantic networks or similar knowledge representation structures still use production rules to do their reasoning.

Production rule systems for diagnostic problem solving are based on techniques for deductive inference. Such systems typically use rules of the form

IF *conjunct of manifestations* THEN *disorder*

to construct chains of deductive reasoning showing that the existence of some set of manifestations provides evidence for the presence of various causes or disorders. Such reasoning proceeds from the rule of modus ponens: if "A implies B" is true and if "A" is true, then "B" is true. However, diagnostic problem solving is basically an abductive process rather than a deductive one. Abductive inference proceeds from a rule which goes in the reverse direction from modus ponens: if "A implies B" is true and "B" is true, then *possibly* "A" is true.

In this paper, we argue that in diagnostic problems where more than one disorder occurs simultaneously, the usual deductive approach used in production rule systems can lead to severe problems. We briefly describe an abductive approach to diagnostic problem solving which overcomes some of these problems, and discuss some knowledge-representation and control issues which arise in the abductive approach.

---

[1] This work was supported in part by Contract NB83SBCA2124 from the National Bureau of Standards, and in part by Contract N60921-83-C-0107 from the Naval Surface Weapons Center.

2. Diagnostic Problem Solving

In order to compare various approaches to diagnostic problem solving, we first need to formalize what is meant by a diagnostic problem. In our formation every diagnostic problem domain has the following characteristics:

1. There are various disorders which can occur in that domain which may or may not be present in specific diagnostic problems. The set of all possible such disorders we call D. For the purposes of this paper, we assume that all members of D are independent of each other.

2. If a disorder d are present, then it may cause one or more symptoms, signs, or *manifestations* of its presence. The set of all possible manifestations in a given problem domain we call M. We let $C \in D \times M$ be the relation between disorders and the manifestations they cause; i.e., $(d, m) \in C$ if and only if d is capable of causing m. In this paper, we assume that a disorder does not necessarily always cause all of the manifestations it is capable of causing, as is often the case in real-world diagnostic problems.

Thus, a problem domain may be specified as a three-tuple $<D, M, C>$.

Within a problem domain, a diagnostic problem occurs when one or more manifestations are present. Thus a diagnostic problem P may be specified as a four-tuple $P = <D, M, C, M^+>$, where $M^+ \subseteq M$ is the set of all manifestations which are actually present. Given P, the task is to find the set of disorders $D^+ \subseteq D$ which is responsible for the presence of the manifestations in M^+.

It may not be possible unambiguously to determine D^+, as there may be more than one set of disorders capable of causing M^+. Several possible criteria have been proposed for how to determine D^+, and some of them are listed below.

Criterion 1. D^+ is the set of all disorders capable of causing any of the manifestations in M^+.

Criterion 2. Every set of disorders capable of causing M^+ is an alternate hypothesis for the identity of D^+.

Criterion 3. Not all of the alternate hypotheses produced by Criterion 2 need be considered. If (as we are assuming) all disorders are independent of each other, then it follows that the simplest possible explanation for a set of manifestations M^+ is a minimum[2] set of disorders capable of causing M^+. From Ockham's razor, it follows that such a set of disorders is likely to be the correct diagnosis. In general, there may be several smallest sets of disorders capable of causing M^+, and according to Criterion 3 these sets are alternate hypotheses for the identity of D^+.

Criterion 4. Although not all of the alternate hypotheses produced by Criterion 2 need be considered, more than just the minimum sets should be considered. For example, suppose that M^+ can be caused either by one very rare disorder d_1, or by two very common disorders d_2 and d_3. Then even though $\{d_1\}$ is a simpler explanation, $\{d_2, d_3\}$ may be more likely and should also be considered as a hypothesis. However, a set such as $\{d_1, d_2\}$ need not be considered as a hypothesis, because if d_1 is present, then it alone is capable of causing all occurring manifestations and so there is no evidence for the additional presence of d_2. As a generalization of this example, the set of possible

[2] i.e., having the smallest possible cardinality

alternatives for D^+ according to Criterion 4 is the set of all minimal[3] sets of disorders capable of causing M^+.

3. An Example

As an example, consider the following problem domain:

disorder d_1 can cause manifestation m_1;
disorder d_2 can cause manifestations m_1 and m_2;
disorder d_3 can cause manifestations m_2, m_3, and m_4;
disorder d_4 can cause manifestation m_4;
there are no other disorders;
all of the disorders are independent of each other;
a disorder does not always cause all of the disorders it is capable of causing.

Three diagnostic problems in this problem domain are discussed below.

Problem P1: suppose that m_1 is present. Then the correct answers according to each of the criteria are

Criterion 1: d_1 & d_2;
Criterion 2: $d_1 \mid d_2 \mid (d_1$ & $d_2)$;
Criterion 3: $d_1 \mid d_2$;
Criterion 4: $d_1 \mid d_2$.

Problem P2: suppose that both m_1 and m_2 are present. Then the correct answers according to each of the criteria are

Criterion 1: d_1 & d_2 & d_3;
Criterion 2: $(d_1$ & $d_3) \mid d_2 \mid (d_1$ & $d_2) \mid (d_2$ & $d_3) \mid (d_1$ & d_2 & $d_3)$;
Criterion 3: d_2;
Criterion 4: $(d_1$ & $d_3) \mid d_2$.

Problem P3: suppose m_1, m_2, m_3, and m_4 are present. Then the correct answers according to each of the criteria are

Criterion 1: d_1 & d_2 & d_3 & d_4;
Criterion 2: $(d_1$ & $d_3) \mid (d_2$ & $d_3) \mid (d_1$ & d_2 & $d_3) \mid (d_1$ & d_3 & $d_4)$
$\mid (d_2$ & d_3 & $d_4) \mid (d_1$ & d_2 & d_3 & $d_4)$;
Criterion 3: $(d_1$ & $d_3) \mid (d_2$ & $d_3)$;
Criterion 4: $(d_1$ & $d_3) \mid (d_2$ & $d_3)$.

Note that in this example, if the answers are considered as Boolean expressions, then in each case the answer produced by Criterion 3 implies the answers produced by Criteria 2 and 4, and the answers produced by Criteria 2 and 4 are equivalent. Although the proof is beyond the scope of this paper, this property can be proved to be true in general.

4. Problems with Production Rule Systems

Let us consider how a production rule system might perform in the above example. If we were to write the problem domain knowledge naively in the form of production

[3] i.e., a set D^+ such that no proper subset of D^+ has the same property.

rules, we might write

R1. IF m_1 THEN d_1 (c_1)
R2. IF m_1 & m_2 THEN d_2 (c_2)
R3. IF m_2 & m_3 & m_4 THEN d_3 (c_3)
R4. IF m_4 THEN d_4 (c_4)

where c_1, c_2, c_3, and c_4 are the relative certainties with which we believe that the disorders d_1, d_2, d_3, and d_4 are present. The results produced by these rules are

d_1 in Problem P1;
d_1 & d_2 in Problem P2;
d_1 & d_2 & d_3 & d_4 in Problem P3.

Only in Problem P3 does the answer exactly fit any of the four criteria. Part of the problem is that not enough rules fire, since each rule requires the presence of all of its preconditions in order to fire. This problem can be handled by adding additional rules referring to all combinations of manifestations which can be caused by each disorder:

R5. IF m_1 THEN d_2 (c_5)
R6. If m_2 THEN d_2 (c_6)
R7. IF m_2 THEN d_3 (c_7)
R8. IF m_3 THEN d_3 (c_8)
R9. IF m_4 THEN d_3 (c_9)
R10. IF m_2 & m_3 THEN d_3 (c_{10})
R11. IF m_2 & m_4 THEN d_3 (c_{11})
R12. IF m_3 & m_4 THEN d_3 (c_{12})

If these rules are added, then the system will produce (with varying certainties)

d_1 & d_2 in Problem P1;
d_1 & d_2 & d_3 in Problem P2;
d_1 & d_2 & d_3 & d_4 in Problem P3.

These results all fit Criterion 1.

It seems fairly clear that the Criterion 1 is not really an adequate characterization of the solution to a diagnostic problem. However, if we are restricted to using rules of the form

IF *manifestations* THEN *disorder*

then the production rule approach cannot be made to satisfy Criteria 2, 3, or 4 because the only conclusions it will be able to produce are conjuncts of disorders.

5. Another Approach

Part of the reason why the usual rule-based approach cannot produce diagnoses satisfying Criteria 2, 3, or 4 is that the information contained in the production rules is simply incorrect. The underlying causal knowledge is not of the form

IF *manifestations* THEN *disorder*

typically found in rule-based expert systems but is instead of the form

IF *disorder* THEN *manifestations.*

Suppose, for example, that a manifestation m_1 can be caused by any of the disorders d_1, d_2, and d_3. If m_1 is present, then we cannot deduce the presence of d_1, nor of d_2, nor of d_3. The correct action would instead be to postulate d_1, d_2, and d_3 as alternate possible hypotheses for what is causing m_1. However, if we further knew that d_1, d_2, and d_3 were the *only* disorders capable of causing m_1, then we could correctly deduce that at least one of d_1, d_2, and d_3 must be present. This is a special case of a diagnostic principle which we will call the *principle of abduction*:

Theorem 1. Suppose that M^+ is a set of manifestations which can be caused by any of the sets of disorders $D_1, D_2, ..., D_k$, and that there are no other sets of disorders capable of causing M^+. Suppose also that the manifestations in M^+ can never occur without being caused. Then if these manifestations are present, then one of the sets of disorders $D_1, D_2, ..., D_k$ must be present.

Proof. Immediate from modus ponens.

The principle of abduction allows us to translate causal knowledge of the form

IF *disorder* THEN *disjunct of manifestations*

into equivalent diagnostic knowledge of the form

IF *manifestation* THEN *disjunct of causes.*

For example, the information about the problem domain used for Problems P1, P2, and P3 can be directly written as a problem in abductive inference with the following set of rules. Note that these rules go in the reverse direction from rules R1 through R4, and that (in terms familiar to the logician) these rules are not Horn clauses.

A1. IF d_1 THEN m_1
A2. IF d_2 THEN $m_1 \mid m_2$
A3. IF d_3 THEN $m_2 \mid m_3 \mid m_4$
A4. IF d_4 THEN m_4

The principle of abduction allows us to translate this abductive problem into a deductive problem having the following set of rules.

D1. IF m_1 THEN $d_1 \mid d_2$
D2. IF m_2 THEN $d_2 \mid d_3$
D3. IF m_3 THEN d_3
D4. IF m_4 THEN $d_3 \mid d_4$

These rules will produce the diagnoses

$d_1 \mid d_2$ for Problem P1;
$(d_1 \mid d_2)$ & $(d_2 \mid d_3)$ for Problem P2;
$(d_1 \mid d_2)$ & $(d_2 \mid d_3)$ & d_3 & $(d_3 \mid d_4)$ for Problem P3.

If considered as Boolean expressions, the answers produced by the rules {D1, D2, D3, and D4} are logically equivalent to the answers produced by Criterion 2 and the answers produced by Criterion 4. Furthermore, they contain all of the disorders appearing in the answers produced by Criterion 1. From this it follows that each answer produced by the rules {D1, D2, D3, D4} can be transformed exactly into the answers produced by Criteria 1, 2, 3, and 4 by means of the following operations:

1. To transform the answer into the one produced by Criterion 1, take the conjunct of all the disorders appearing in the answer.
2. To transform the answer into the one produced by Criterion 2, treat it as a Boolean expression and find all implicants.

3. To transform the answer into the one produced by Criterion 3, treat it as a set covering problem and find all minimum set covers.
4. To transform the answer into the one produced by Criterion 4, treat it as a Boolean expression and find all prime implicants.

6. Control Strategies

The previous section described a general approach to diagnostic problem solving. The main steps in the approach are as follows:

Step 1. Formulate the diagnostic problem as an abductive problem using the direct causal knowledge about the problem. This results in a set of rules similar to rules A1 through A5.

Step 2. Use the Principle of Abduction to translate the abductive problem into an equivalent deductive problem. This results in a set of rules similar to rules D1 through D5.

Step 3. Apply the deductive rules to the problem.

Step 4. Use one of the transformational techniques listed above to find an answer satisfying Criteria 1, 2, 3, or 4 as desired.

In implementing this approach, a number of problems must be solved. Five major ones are described below.

1. Diagnostic problem solving normally requires sequential hypothesize-and-test approach. Typically, only a few of the members of M^+ will be known to be present at the outset. The diagnostician will have a tentative hypothesis sufficient to explain the members of M^+ that are known to be present, and based on this hypothesis he or she will perform tests or ask questions to discover additional members of M^+. As these new manifestations are discovered, the hypothesis is revised to account for them. In order to be broadly useful, knowledge-based diagnostic problem solving systems must perform in a similar way.
2. Although our approach has been described in terms of four separate steps, it is often more efficient--particularly in the case of Criteria 3 and 4--to write a procedure which solves the problem directly as a problem in abductive inference.
3. While solving large diagnostic problems, the alternate hypotheses for sets of disorders capable of causing M^+ can become quite large and unwieldy. A representation is needed which is compact, efficient, and easily understandable.
4. Although this paper has so far ignored the issues raised by the use of certainty factors in diagnostic problem solving, a way is needed to handle them.
5. The rule format

   IF *conjunct of manifestations* THEN *disorder*,

   is the standard form for production rules in knowledge-based computer systems. If (as this paper proposes) we instead use rule formats such

   IF *disorder* THEN *disjunct of manifestations*.

   or

   IF *manifestation* THEN *disjunct of disorders*,

   then it is no longer clear how to do rule chaining.

We have developed an algorithm for diagnostic problem solving using Criterion 3 which handles the first four problems described above [7] [8], and which has been successfully used as the control strategy for knowledge-based diagnostic problem solvers in several different domains [9]. This algorithm is summarized below; for a more detailed treatment, the reader is referred to [7] [8]. We are currently extending the algorithm to handle the fifth problem (how to do inferential chaining) [6], and to work for Criterion 4 as well as Criterion 3 [11].

Three main data structures are used in the algorithm:

1. MANIFS $\subseteq M^+$ is the set of manifestations known to be present so far, i.e., our current hypothesis for the identity of M^+.
2. SCOPE $\subseteq D$ contains every disorder capable of causing at least one of the manifestations in MANIFS (note that this is the answer Criterion 1 would produce).
3. FOCUS is the family of all minimum sets of disorders capable of causing MANIFS. This is the set of alternate hypotheses for D^+ according to Criterion 3. FOCUS is also our solution to the problem of how to represent the alternate hypotheses compactly: it is expressed as a disjunct of conjuncts of disjuncts, and is manipulated directly in that form.

At the top level, the algorithm is a hypothesize-and-test loop which looks roughly as follows:

```
procedure HT
    MANIFS := SCOPE := FOCUS := ∅
    while not all of M⁺ is known do
        perform a test to discover a new manifestation m ∈ M⁺
        MANIFS := MANIFS ∪ {m}
        SCOPE := SCOPE ∪ {d ∈ D | (d,m) ∈ C}
        adjust FOCUS to accomodate m
    endwhile
    return FOCUS
end HT
```

The knowledge base for this procedure is a set of frames, one for each disorder $d \in D$. These frames are generalizations of abductive rules such as the rules {A1, A2, A3, A4} discussed earlier. The frame for a particular disorder d contains all the information we may have about d—what manifestations it is capable of causing, what conditions govern whether it may occur, etc. These frames thus define the causal relation C. In addition, the frames contain information about how likely it is to cause each of its manifestations, and this information is used to determine the relative likelihoods of the various alternate hypotheses. This provides a straightforward solution to the problem of how to handle certainty factors.

In adjusting FOCUS to accomodate a new manifestation m, there are two possible cases which may occur:

1. Some of the sets of disorders in FOCUS may be capable of causing MANIFS $\cup \{m\}$. In this case, it can be proved that the family of minimum covers capable of causing MANIFS $\cup \{m\}$ is a subfamily of FOCUS. Thus FOCUS must be adjusted to remove the sets of disorders that no longer work.
2. MANIFS $\cup \{m\}$ cannot be caused by any set of disorders in FOCUS. In this case, it can be proved that each minimum set of disorders capable of causing

MANIFS $\cup \{m\}$ has cardinality exactly one more than the cardinality of each minimum set of disorders capable of causing MANIFS. In this case, FOCUS must be completely recomputed, but the problem is not as difficult as it might be in general since we know exactly how many disorders will appear in each set.

7. Another Example

We now illustrate the operation of the algorithm on Problem P3. Just how tests are generated in line 3 is described in [9]. The final result produced is independent of the order in which these manifestations are discovered, so for now, let us assume the members of M^+ are discovered in the order m_1, m_2, m_3, m_4. Then the following events occur.

1. Initially, MANIFS = SCOPE = FOCUS = $\emptyset$.

2. m_1 is found to be present. Then MANIFS is set to $\{m_1\}$. SCOPE is set to $\{d_1,d_2\}$, the set of all disorders which can cause m_1. MANIFS can be caused by d_1 alone or d_2 alone, so FOCUS is the expression $d_1|d_2$.

   If we were solving Problem P1, then at this point no more manifestations would be found and the program would terminate, returning FOCUS. This would be the correct answer to P1 according to Criterion 3.

3. m_2 is found to be present. The disorders capable of causing m_2 are d_2 and d_3. These are added into SCOPE, yielding SCOPE $= \{d_1,d_2,d_3\}$. MANIFS can still be caused by d_2, but not by d_1. Thus FOCUS is the expression d_1.

4. m_3 is found to be present. The only disorder capable of causing m_3 is d_3; thus SCOPE does not change. No single disorder can now explain MANIFS. Thus FOCUS is recomputed, looking for sets of two disorders each. Both $\{d_1,d_3\}$ and $\{d_2,d_3\}$ work, and thus FOCUS is represented as the expression $(d_1 \mid d_2)$ & d_3.

   If we were solving Problem P2, then at this point no more manifestations would be found and the program would terminate, returning FOCUS. This would be the correct answer to P2 according to Criterion 3.

5. m_4 is found to be present. m_4 can be caused by either d_3 or d_4, and thus SCOPE becomes $\{d_1,d_2,d_3,d_4\}$. m_4 can be caused by both of the sets of disorders represented in the previous FOCUS, so FOCUS does not change.

6. No more manifestations are found. Thus the program terminates and returns FOCUS, which is the correct answer to Problem 3 according to Criterion 3.

8. Conclusions

We have discussed four different criteria for what constitutes the solution to a diagnostic problem. The ordinary deductive production rule approach to diagnostic problem solving uses rules of the form

IF *conjunct of manifestations* THEN *disorder*,

and this approach is sufficient to meet only the least sophisticated of these criteria.

Part of the reason for this problem is that diagnostic problem solving is more properly an abductive problem rather than a deductive one: the causal knowledge we have about diagnostic problems is not of the form given above, but rather is of the form

IF *disorder* THEN *disjunct of manifestations.*

The principle of abduction stated in this paper shows how to translate this kind of knowledge into knowledge that can be used in deductive problem solving. This results in rules of the form

IF *manifestation* THEN *disjunct of disorders.*

Implementing a rule-based problem solver using this kind of knowledge requires a much different kind of control strategy than is normally used in production systems. We have discussed the basics of a control strategy for this approach.

Production rules have been criticized in the past as a representation of diagnostic knowledge [10], and our approach appears in at least some cases to be a more intuitively plausible descriptive representational formalism and model of diagnostic reasoning. A knowledge-based reasoning system based on this approach has been implemented and successfully used for several diagnostic medical problems. [9].

The inference method used in INTERNIST [2] differs but is rather close to our approach [8] and some similar techniques have also been used on an immunological problem [4] [12]. However, this approach has not to our knowledge been previously examined in detail as a model of general diagnostic reasoning.

Our approach also appears to have application (perhaps in modified form) to other types of problems as well. It has been used for several non-medical "toy" diagnostic problems, and an effort is underway to use it for a non-diagnostic problem in automated manufacturing [5].

9. References

[1] Duda, R. O., Hart, P. E., Barrett, P., Gaschnig, J. G., Konolige, K., Reboh, R., and Slocum, J., Development of the Prospector Consultation System for Mineral Exploration, Tech. Report, SRI International, Menlo Park, CA, Oct. 1978.

[2] Miller, R., Pople, H., and Myers, J., Internist-1: An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine, *NEJM* **307**, pp. 468-476, 1982.

[3] Nau, D. S., Expert Computer Systems, *Computer* **16**, 2, pp. 63-85, Feb. 1983.

[4] Nau, D. S., Markowsky, G., Woodbury, M. A., and Amos, D. B., A Mathematical Analysis of HLA Serology, *Mathematical Biosciences* **40**, pp. 243-270, 1978. (An early version is available as Tech. Report RC 6803, IBM T.J. Watson Research Center, Yorktown Heights, NY, Oct. 1977.)

[5] Nau, D. S., Reggia, J. A., Blanks, M. W., Peng, Y., and Sutton, D., Artificial Intelligence Approaches for Automated Process Planning and Control, Tech. Report, Computer Science Dept., University of Maryland, College Park, Md., Feb. 1984. Submitted for publication.

[6] Peng, Y., Ph.D Dissertation, in progress. 1984.

[7] Reggia, J. A., Nau, D. S., and Wang, P. Y., A Formal Model of Diagnostic Inference, Part One: Problem Formulation and Decomposition, submitted for publication, 1983.

[8] Reggia, J. A., Nau, D. S., and Wang, P. Y., A Formal Model of Diagnostic Inference, Part Two: Algorithmic Solution and Applications, submitted for publication, 1983.

[9] Reggia, J. A., Nau, D. S., and Wang, P. Y., Diagnostic Expert Systems Based on a Set Covering Model, *International Journal of Man-Machine Studies*, pp. 437-460, Nov. 1983.

[10] Reggia, J. A., Nau, D. S., and Wang, P. Y., Diagnostic Expert Systems Based on a Set Covering Model, To appear. Nov. 1983.

[11] Reggia, J. A., Peng, Y., and Nau, D. S., A Formal Model of Diagnostic Inference, Part Three: Exploring the Nature of Parsimony, In preparation, 1984.

[12] Woodbury, M., Ciftan, E., and Amos, D., HLA Serum Screening Based on an Heuristic Solution to the Set Cover Problem, *Comp. Pgm. Biomed.* **9**, pp. 263-273, 1979.