Dana

Preprint:
WORKSHOP ON NON-MONOTONIC
REASONING
Oct. 1984

AN ABDUCTIVE NON-MONOTONIC LOGIC

James A. Reggia
Dana S. Nau

Department of Computer Science
University of Maryland
College Park, MD 20742
USA

Diagnostic Problems and Abductive Inference

In the past, most work on automated reasoning with plausible or default assumptions has concentrated on extending current methods for **deductive** inference (e.g., [1], [2], [3], [4], [5]). While this research has in many ways been quite successful, it has proven very difficult to adopt for solving a number of real-world problems that involve non-monotonic reasoning.

As an example of these difficulties, consider a diagnostic problem where one is given a set of manifestations and must explain why they are occurring by postulating the presence of one or more causative disorders. Problems of this kind are very common: they include diagnosing a patient's signs and symptoms in medicine, determing why a computer program failed, deciding why an automobile will not start, finding the cause of noises in a plumbing system, localizing a fault in an electronic circuit, explaining why a child makes arithmetic mistakes, etc. Diagnostic problem solving inherently involves many features that have been said to characterize non-monotonic reasoning [2]. For example, during problem solving one must handle ambiguity (alternative causative disorders under consideration), use of causal associations (disorder d causes manifestation m), making inferences based on incomplete information (the diagnostician starts with one or more initial manifestations and formulates a tentative diagnosis), revision of beliefs in the face of additional information (real-world diagnostic problem solving is typically a sequential, hypothesize-and-test process [6]), etc.

Unlike deductive reasoning, which in its simplest form might be characterized by modus ponens

Given Fact "A" and Rule "$A \longrightarrow B$", infer "B",

diagnostic problem solving inherently involves **abductive** inferences of the form

Given Fact "B" and Association "$A \longrightarrow B$", infer "plausible A".

Although the "$\longrightarrow$" in the deductive syllogism refers to logical implication, in the abductive syllogism it refers to a causal association between A and B: "disorder A is capable of causing symptom B, and symptom B is known to be present, so perhaps disorder A is causing it."

Our approach to modeling abductive reasoning has been to develop a simple but formal abductive logic referred to as the generalized set covering (GSC) model [7]. This non-deductive theoretical model has been used as the basis of a number of implemented diagnostic expert systems [8], and has proven quite powerful: it supports a descriptive knowledge representation and answer justification [9], and it avoids many of the difficulties that occur with deductive reasoning in the context of multiple simultaneous disorders. The GSC model is briefly summarized in the next section, and a discussion of its inherently non-monotonic nature is then provided. A detailed discussion of the relationships between deductive and abductive inference is available in [10].

The GSC Model

In the GSC model, a **diagnostic problem** P is defined to be a 4-tuple $P = \langle D,M,C,M^+\rangle$. There are two discrete finite sets which characterize the scope of the diagnostic problem: D, representing all possible **disorders** d_i that can occur, and M, representing all possible **manifestations** m_j (symptoms) that may occur when one or more disorders are present. The relation $C \subseteq D \times M$ captures the intuitive notion of causation, where $\langle d_i, m_j \rangle \in C$ represents "d_i can cause m_j" (this is the basis of the abductive syllogism described earlier). $M^+ \subseteq M$ represents those manifestations present in a specific problem.

Within this framework, the non-empty set of all possible manifestations of any disorder d_i is designated

$$\text{man}(d_i) = \{m_j \mid \langle d_i, m_j \rangle \in C\}$$

and the non-empty set of disorders which can cause any manifestation m_j as

$$\text{causes}(m_j) = \{d_i \mid \langle d_i, m_j \rangle \in C\}.$$

This concept can be generalized to a set of disorders D_I and a set of manifestations M_J as

$$\text{man}(D_I) = \bigcup_{d_i \in D_I} \text{man}(d_i) \qquad \text{and} \qquad \text{causes}(M_J) = \bigcup_{m_j \in M_J} \text{causes}(m_j).$$

Also, $\text{man}^+(d_i)$ is used as an abbreviation for $\text{man}(d) \wedge M^+$.

Given $P = \langle D, M, C, M^+ \rangle$, $E \subseteq D$ is defined to be an **explanation** for M^+ if

(i) $M^+ \subseteq \text{man}(E)$, or in words: E **covers** M^+; and

(ii) E is **parsimonious**.

This definition captures many features of what one means by "explaining" a set of manifestations. Part (i) specifies the reasonable constraint that an explanation E must be able to cause all manifestations known to be present in the case being diagnosed. The "reasonableness" of this first constraint comes from the following theorem (from [10]):

Theorem (Principle of Abduction for Diagnostic Problems). Suppose that M^+ is a set of manifestations which can be caused by any of the sets of disorders $D_1, D_2, \ldots D_N$, and 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, one of the sets of disorders $D_1, D_2, \ldots D_N$ must be present.

Part (ii) specifies that E must also be "parsimonious," reflecting "Ockham's Razor": the simplest explanation is the preferable one. Thus, given the Principle of Abduction, the real issue raised by the definition of an explanation is how one should go about formalizing the notion of parsimony or simplicity. For example, if one replaces (ii) in the definition of an explanation with

$$\text{(ii')} \quad |E| \leq |D| \text{ for any other cover } D \text{ of } M^+$$

then one has equated simplicity with minimal cardinality, and an explanation with a minimal cover [11].

While this is a very plausible concept of explanation in some situations, in others it is inadequate. For example, two very common disorders might be a more plausible explanation than a single rare disorder in some applications [12]. In such situations, (ii) in the definition of an explanation might be replaced with

$$\text{(ii")} \quad \text{no proper subset of } E \text{ covers } M^+$$

equating simplicity with irredundancy (E contains no "redundant" disorders) and an explanation with an irredundant cover.

Other notions of parsimony are possible, and the interested reader is referred to the appropriate references [10,12,13]. For concreteness, in this paper we will restrict our attention to parsimony as defined in (ii').

In general diagnostic problem solving, for any given manifestation there may be more than one disorder which can cause that manifestation, reflecting the interpretive ambiguity of manifestations as diagnostic clues. The diagnostician is therefore generally interested in identifying <u>all</u> plausible explanations which might be present, at least during the early stages of problem solving. Thus, in the GSC model, the **solution** to a diagnostic problem is defined to be the set of all explanations for M^+. Rather than represent the solution to a diagnostic problem as a simple list of explanations, it turns out to be advantageous (see [7]) to represent these explanations as a set of **generators**. A generator is a collection of sets of "competing" disorders that implicitly represents a set of explanations in the solution and can be used to generate them. A generator is analogous to a Cartesian product, the difference being that the generator produces unordered sets rather than ordered tuples. For example, the generator $(d_1,d_2) \times (d_3,d_4) \times (d_5,d_6,d_7)$ represents 12 explanations: $\{d_1,d_3,d_5\}$, $\{d_1,d_3,d_6\}$, etc. (We use round parentheses to designate sets of disorders in generators for clarity). More than one generator may be needed to represent the solution to a diagnostic problem.

Given just the above formulation, it is possible to prove a number of interesting properties of diagnostic problems [7]. Criteria for considering disorders to be "competitors" or alternatives to one another can be stated. For example (assuming (i) and (ii')), the following holds:

Theorem (Competing Disorders Theorem). Let E be an explanation for M^+, and let $man^+(d_1) \subseteq man^+(d_2)$ for some d_1, $d_2 \in D$. Then:

(a) d_1 and d_2 are not both in E,

(b) if $d_1 \in E$, then $E' = (E-\{d_1\}) \cup \{d_2\}$ is also an explanation for M^+.

As another example, criteria can be given for decomposing diagnostic problems into independent subproblems based on a notion of connectedness of manifestations [7].

We have presented in detail elsewhere a formal algorithm for solving sequential diagnostic problems, and proven its correctness [7]. This algorithm (called HT, for "hypothesize-and-test") is initially given a (usually incomplete) set of manifestations which are known to be present, and derives a set of generators representing a **tentative** solution. This tentative hypothesis is then used to guide the "question generation" process which uncovers additional manifestations until M^+ is completely discovered. During this process the hypothesized solution is repeatedly modified in the light of new information so that it continues to represent all explanations for the known manifestations.

Algorithm HT uses three principle data structures:

1. MANIFS $\subseteq M^+$, the set of manifestations known to be present so far;
2. SCOPE $\subseteq D$, the set causes(MANIFS); and
3. HYPOTHESIS, a set of generators representing the tentative solution for MANIFS with the assumption (usually incorrect) that MANIFS = M^+.

Roughly, the top-level structure of algorithm HT is:

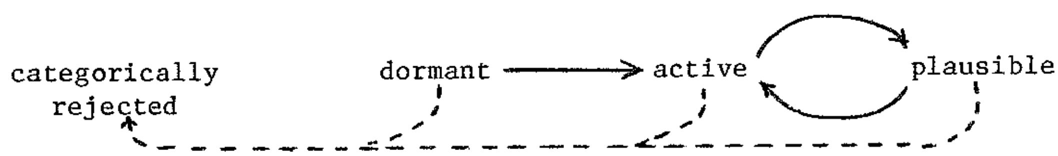
```
function HT
  MANIFS := SCOPE: = Ø;
  HYPOTHESIS := {Ø};
  while not all of M+ is known do
    discover a new manifestation m ε M+;
    MANIFS := MANIFS v {m};
    SCOPE: := SCOPE v causes(m);
    adjust HYPOTHESIS to accomodate m
  endwhile
  return HYPOTHESIS
end HT.
```

The adjustment of the HYPOTHESIS is based on set operations involving causes(m) and the sets in the generators in HYPOTHESIS [7]. In adopting HT to real-world problem solving, many issues must be addressed: question generation to guide manifestation discovery, termination criteria, detection and handling of "unexplainable manifestations," incorporation of probabilistic information, etc. Our approach to these issues is described in [8].

Non-Monotonic Aspects of the GSC Model

As implied earlier, the GSC model inherently incorporates a number of features characteristic of non-monotonic logics [2]: ambiguity, use of causal associations, the closed-world assumption, making inferences in the context of incomplete information with subsequent revision of beliefs, etc. Perhaps the most striking feature of the GSC model when compared to deductive logics is that the derivation of individual "theorems" (by which we mean inferences about the presence or absence of individual disorders) at any point in problem solving is a global property of available information. Unlike deductive approaches where "local" inferences like "$fact_1$ + rule $\Longrightarrow$ $fact_2$" or "$clause_1$ + $clause_2$ $\Longrightarrow$ resolvent" occur, in the abductive GSC model a "fact + association" does not in isolation result in a theorem.

During problem solving, the logical status of each possible $d_i \in D$ dynamically changes between four mutually exclusive states as illustrated below:



The left-to-right ordering of the logical states above is meant to indicate a scale from "false" (leftmost) to "presumed true" (rightmost). The status of any disorder in the context this scale represents the abductive equivalent of theorems in deductive logics. Permissible transitions in logical status are indicated by arrows. Initially, every disorder d_i is dormant, and it becomes active only when some $m_j \in man(d_i)$ is discovered to be present. If a disorder is part of the current hypothesized solution, it becomes "plausible" (i.e., its possible presence is inferred). As subsequent manifestations are discovered, the logical status of d_i may oscillate between being active and plausible, representing a revision of belief in the face of additional information. This non-monotonic behavior, represented by the paired, oppositely directed arrows above, is even more striking in that the very form of the hypothesis or "beliefs" may change during problem solving. For example, at one point it may be believed that a single disorder is present (which disorder it is may not be certain), but as additional information is uncovered this belief may be revised to reflect that two disorders must be present. All of these inferences occur "automatically" and are a consequence of the basic definitions of the GSC model.

As a simple example, suppose that a diagnostic problem $P = \langle D,M,C,?\rangle$ is given where $D = \{d_1,d_2,d_3\}$, $M = \{m_1,m_2,m_3,m_4,m_5\}$, and C is implicitly specified by knowing that $man(d_1) = \{m_1,m_3\}$, $man(d_2) = \{m_3,m_4\}$, and $man(d_3) = \{m_2,m_4,m_5\}$. Initially all $d_i \in D$ are dormant. Suppose algorithm HT first discovers that m_3 is present. Then it would construct SCOPE = $\{d_1,d_2\}$ and the initial HYPOTHESIS would be the generator (d_1,d_2), representing the belief that either $\{d_1\}$ or $\{d_2\}$ is a plausible explanation for the known manifestation. At this point in problem solving, both d_1 and d_2 would be plausible, and d_3 would be dormant.

If HT next discovered that m_4 is present, it would form SCOPE = $\{d_1,d_2,d_3\}$ and the HYPOTHESIS would be the generator (d_2), which generates the single explanation $\{d_2\}$. At this point, only d_2 is plausible, while d_1 and d_3 are active but not plausible (recall that we are using criteria (ii') as our notion of parsimony). Finally, if HT then discovered that m_1 is present, the SCOPE would be unchanged while the HYPOTHESIS would be represented by the generator $(d_1) \times (d_2,d_3)$. All three disorders would be plausible now, and if no further manifestations occurred, it could be said that algorithm HT had inferred the plausible presence of d_1 <u>and</u> one of d_2 or d_3.

We do not discuss "categorical rejection," the fourth possible logical state of disorders here (see [7,8]), but only indicate that it provides one example of **deductive** abilities that can be captured within the framework of the GSC model.

Conclusion

In this paper we have indicated that diagnostic problem solving involves many aspects of non-monotonic reasoning, and we have described the GSC model of diagnostic inference and its non-monotonic features. Perhaps surprisingly, many real-world problems that upon superficial examination do not appear to be "diagnostic problems" per se seem amenable to solution with abductive inference using the GSC model. For example, we have studied the use of our model for selection of which statistical tests to use for data analysis [14], selection of machining operations during process planning in manufacturing [15], and construction of models of nerve cell growth [16]. The common factor in all of these application areas seems to be the need to select a parsimonious set of choices from alternatives. For that reason, we believe that many of the basic concepts introduced in the GSC model go beyond diagnostic inference and apply to common sense reasoning in a much broader context.

Our work on abductive inference in general and the GSC model in particular has just scratched the surface of what we believe is an important aspect of non-monotonic reasoning. We are currently studying the extension of the GSC model to a much broader range of issues: support of answer justification [9], non-minimal covers, non-independent disorders, causal chaining, incorporation of probabilistic information, relationship to deductive inference [10], etc. The interested reader is referred to [13] for details of these issues. We believe that this approach to non-monotonic inference holds great promise for the future, both from a theoretical and a practical point of view.

References

[1] McDermott D and Doyle J: Non-Monotonic Logic I, Art. Intell., 13, 1980, 41-72.

[2] Reiter R: On Reasoning by Default, Proc. TINLAP II, 1978, 210-218.

[3] Doyle J: A Truth Maintenance System, Art. Intell., 12, 1979, 231-272.

[4] Sridharan N, Schmidt C, and Goodson J: Reasoning by Default, CBM-TR-119, Rutgers University, 1981.

[5] McDermott D: Assimilation of New Information by a Natural Language-Understanding System, AI TR-291, MIT AI Laboratory, Feb. 1974.

[6] Reggia J: Computer-Assisted Medical Decision Making, in Applications of Computers in Medicine, M. Schwartz (ed.), IEEE Press, 1982, 198-213.

[7] Reggia J, Nau D and Wang P: A Theory of Abductive Inference in Diagnostic Expert Systems, Parts I and II, TR-1338, Department of Computer Science, University of Maryland, December 1983, submitted for publication.

[8] Reggia J, Nau D and Wang P: Diagnostic Expert Systems Based on a Set Covering Model, Int. J. Man-Machine Studies, 19, 1983, 437-460.

[9] Reggia J, Perricone B, Nau D and Peng Y: Answer Justification in Abductive Inference Systems, IEEE Trans. Biomed. Engr., in press.

[10] Nau D and Reggia J: Relationships Between Deductive and Abductive Inference in Knowledge-Based Diagnostic Problem Solving, submitted for publication, 1984.

[11] Edwards J: Coverings and Packings in a Family of Sets, Bull. Ann. Math. Soc., 68, 1962, 494-499.

[12] Reggia J, Nau D, and Peng Y: A Theory of Abductive Inference in Diagnostic Expert Systems, Part III, manuscript in preparation, 1984.

[13] Peng Y: A General Theoretical Model for Abductive Diagnostic Expert System, TR-1402, Dept. of Computer Science, University of Maryland, May 1984.

[14] Reggia J: Knowledge-Based Decision Support Systems: Development Through KMS, TR-1121, Dept. of Computer Science, University of Maryland, Oct. 1981.

[15] Nau D, et al: Artificial Intelligence Approaches for Automated Process Planning and Control, TR-1382, Dept. of Computer Science, University of Maryland, March 1984.

[16] Tagamets M: Abductive Hypothesis Formation and Refinement During Construction of Natural System Models, M.A. Thesis, Applied Mathematics, University of Maryland, College Park, August 1984.