Experience-Aided Diagnosis for Complex Devices

Michel P. Féret and Janice I. Glasgow
Department of Computing & Information Science,
Queen's University, Kingston,
Ontario, Canada, K7L 3N6*
{feret,janice}@qucis.queensu.ca

Abstract

This paper presents a novel approach to diagnosis which addresses the two problems - computational complexity of abduction and device models - that have prevented model-based diagnostic techniques from being widely used. The Experience-Aided Diagnosis (EAD) model is defined that combines deduction to rule out hypotheses, abduction to generate hypotheses and induction to recall past experiences and account for potential errors in the device models. A detailed analysis of the relationship between case-based reasoning and induction is also provided. The EAD model yields a practical method for solving hard diagnostic problems and provides a theoretical basis for overcoming the problem of partially incorrect device models.

Introduction

The diagnostic process, either human or computational, is partly one of abduction. Abduction is a form of inference where from the logical implication $P \rightarrow Q$, and the conclusion Q, one *abductively* infers the antecedent P. Abduction is also described as the process of making conjectures about observable facts that "explain" the facts in a certain way (Peirce 1955).

Diagnostic reasoning also involves deduction. Deduction is used to rule-out some components as potential diagnoses. For example, if the failure of component X always causes symptom Y, and if Y is not present in the set of symptoms (and not hidden by other symptoms), it can be deduced that X is not the cause of the current problem. Bylander *et al.* called for this type of reasoning to reduce the complexity of abduction problems (Bylander *et al.* 1991).

In this paper, we argue that diagnostic reasoning can also benefit from induction by using past experience to derive diagnoses and explanations that are overlooked by the abductive process or incorrectly dismissed by the deductive process mentioned above. Induction is broadly defined as the ability to generalize from examples. In the diagnostic domain, induction is the process by which a past session's symptoms are judged similar enough to those of the current situation to allow the diagnoser to use a past session's diagnosis to reach a diagnosis for the current situation.

We first describe the background of this research in model-based diagnosis (MBD) and case-based reasoning (CBR). The Experience-Aided Diagnosis (EAD) model is then introduced and used to characterize diagnostic errors that can occur from errors present in the device models. The paper concludes by summarizing the contributions of this paper to the fields of MBD and CBR.

Background

MBD has emerged as a relevant research topic from the problems found in traditional ruled-based diagnostic expert systems. These earlier systems were unable to handle unpredicted faults, had poor explanation facilities and did not take advantage of existing design specifications. MBD addresses these problems by using device models containing "deep" and first-principle knowledge which describe the correct, expected behavior of the device. The search for components in an abnormal state is guided by the discrepancies between what is predicted by the model and what is observed in the device. A diagnostic session is triggered when initial symptoms do not match with the predictions of the model. Some models are fault models; they only describe conditions under which a component or a group of components is or might be faulty.

The main task of diagnosis is to find *explanations* for a set of given symptoms. In the context of diagnosis, explanations are conjectures that must either be consistent with the symptoms and the model of the device or entail the symptoms. Console *et al.* define a model for diagnosis that encompasses explanations that satisfy both of these criteria (Console, Theseider Dupre, & Torasso 1989). This model subsumes Reiter's (Reiter 1987) and De Kleer's (de Kleer & Williams 1987) seminal work. Reiter describes an algorithm that com-

*This research was supported through a contract from the Canadian Space Agency (STEAR program), a scholarship and an operating grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada. We also would like to thank Spectrum Engineering Corporation Ltd.,Peterborough, Ontario, Canada.

putes the minimal conflict sets and derives the minimal diagnoses from them. This algorithm is NP-hard in the general case, in agreement with Bylander *et al.*'s results (Bylander *et al.* 1991). Reiter's characterization of diagnoses also corresponds to deKleer and Williams' definition of diagnosis (de Kleer & Williams 1987). Like Reiter, they use the notion of minimal conflict set, computed with TMSs to derive minimal diagnoses.

One problem with MBD is its inherent computational complexity. Abduction in MBD is, in general, NP-hard, making diagnosis untractable for even medium-sized devices. Many researchers have tried to focus the search for (minimal) diagnoses in MBD (Console, Portinale, & Theseider Dupre 1991; de Kleer 1989; 1990; 1992) or to reduce the complexity of MBD methods (e.g. (Mozetic 1990; Friedrich 1992)). Console *et al.* use compiled knowledge to focus the abductive search with necessary conditions. This is equivalent to reducing the size of the search space for minimal conflicts by compiling some deep knowledge. De Kleer has tried to focus the truth maintenance systems either with statistical information (de Kleer 1990), or failure modes (de Kleer 1989) or more refined look-ahead search strategies (de Kleer 1992). Friedrich (Friedrich 1992) uses hierarchy decompositions of the device to improve the efficiency of his diagnostic algorithm (Friedrich 1992). Mozetic defines abstraction operators on the device models, which lead to hierarchical diagnosis as well (Mozetic 1990).

Another problem with MBD is the difficulty of representing complex device models where possible interactions among components are sometimes overlooked. This problem does not arise with logical circuits because the components of such devices are well described, and their connections and interactions are well-understood. Models for complex devices, however, are *not* used for diagnosis in the real-world. Models are also complex in themselves and need validation and verification. Validating and verifying knowledge bases is difficult and little success has been achieved in this area. It is therefore likely that any MBD system will use an incomplete or incorrect device model. The current literature does not account for this extraneous difficulty and is therefore unrealistic.

We next analyze the relationship between CBR and induction and formalize the notion of abductive induction as a means for addressing some of the outstanding problems in MBD.

CBR and Induction

We propose to apply case-based reasoning (CBR) as a means for incorporating experience in the diagnostic process. The philosophy behind CBR is that "raw", unabstracted experiences can be used as a source of knowledge for problem-solving (Riesbeck & Schank 1989). A CBR system stores past experiences in the form of cases. When a new problem arises, the system retrieves the cases most similar to the current problem, then combines and adapts them to derive and criticize a solution. If the solution is not satisfactory, new cases are retrieved to further adapt it to new constraints, expressed from the non-satisfactory parts of the proposed solution. The process is iterated until the proposed solution is judged acceptable. This process is modeled in (Féret 1993).

Induction provides the ability to generalize from examples. When past cases are judged similar enough to the current situation, an implicit generalization is formed as a set containing the case representing the current situation and the past cases used to solve the current problem. These cases are assumed to yield similar solutions. They all belong to an implicit set: the set of cases that are relevant to the current situation. The relationship between CBR and induction is therefore that they are both concerned with generalization from examples, and that the generalizations are built under uncertainty.

There are differences between CBR and induction. First, the generalizations produced by induction are explicit and defined intentionally. There must be a language in which the generalization is expressed, which might differ from the language in which the data or the background theory are expressed. This need for a generalization language, and the existence of many possible generalizations, lead to biases in the produced generalizations. CBR generalizes implicitly. A generalization is expressed by the cases that are judged similar enough to the current situation. There is no need for a representation language, no arbitrary bias and no added computational cost associated with generating the description of the generalization. Second, an inductive conclusion ϕ must entail the data Δ whereas CBR generalizations do not necessarily relate logically to the cases. There is no logical foundation to CBR systems, and some guess work is usually involved in the similarity measures. CBR corresponds more to abductive induction (Peirce 1955), where some uncertainty is involved in the building of the inductive generalizations. Third, induction assumes the existence of a background theory Γ. Most CBR system do not rely on such a formal background. By incorporating CBR into MBD, the CBR component of the system gains the background theory that the device model provides.

We define abductively inductive hypotheses as follows: Given a background theory Γ and a set of data Δ, such that $\Gamma \not\vdash \Delta$, a sentence ϕ is an abductively inductive hypothesis if and only if there exist Δ^+ and Δ^- such that:

- $\Delta = \Delta^+ \cup \Delta^-$, $\Delta^+ \cap \Delta^- = \{\}$,
- $\Gamma \cup \Delta^- \not\vdash \neg\phi$, i.e. ϕ is consistent with the background theory and some of the data Δ^-,
- $\Gamma \cup \{\phi\} \mid\!\sim_{pl} \Delta^+$, i.e. there is some confidence pl that ϕ is the cause of part of the data Δ^+ (denoted by $\mid\!\sim_{pl}$).

Note that the difference between an abductively inductive hypothesis and an inductive diagnostic explanation lies in the difference in modeling causality. An abductively inductive hypothesis is only *believed*, with some level of confidence *pl*, to be the cause of the symptoms in Δ^+, while an inductive diagnostic explanation entails the symptoms in Δ^+.

The partial fault model given in Figure 1 and extracted from (Console, Portinale, & Theseider Dupre 1991) provides examples of the use of inductive diagnoses. Suppose that the diagnoser using this model has already been through Session 1 and that the current diagnostic session is Session 2. The difference between the two Sessions is that Session 2 has less evidence available: $acceleration(irregular)$ is a symptom present in Session 1 but absent in Session 2. By analogy with Session 1, we could reasonably guess that $E_2 = E_1 = \{oil\_cup\_holed\}$. Because this reasoning is by analogy, Δ^+ cannot be entailed by $\Gamma \cup \{\phi\}$ as in traditional induction, and as in Console *et al.*'s definition of abductive diagnoses. This entailment relationship is instead replaced by a belief that $\Gamma \cup \{\phi\}$ explains ψ^+.

We next present the EAD model for diagnosis that addresses the problems of computational complexity and of imperfect device models.

The EAD Model

The practical and theoretical limits of MBD systems was previously discussed. This section presents the EAD model for diagnosis which explicitly combines deduction, abduction and induction to overcome the problems mentioned above. In this model, the role of deduction is made explicit in trying to reduce the size of the search space for explanatory hypotheses. Abductive induction, implemented with a CBR system uses experience to correct diagnostic errors due to errors present in the device models. We illustrate the EAD model using examples based on the device model of Figure 1.

In EAD, a diagnostic problem is characterized by:

- $MODEL$, a device description, a finite set of first-order logic sentences, and a finite set $COMPONENTS$ of constants denoting components.
- CXT, a set of ground context terms denoting a set of contextual data,
- OBS, a set of ground observation terms, denoting the set of observations to be explained, e.g. $m(x)$ where m is a predicate and x is an instantiated variable,
- $\psi^+ \subseteq OBS$,
- $\psi^- = \{\neg m(x)/m(y) \in OBS$, for each admissible instance $m(x)$ of m other than $m(y)\}$
- *rok*, a map from OBS to $COMPONENTS$, denoting the rule-out knowledge,
- Δ, a case base containing past failure diagnostic sessions,
- *cbr*, a map from $OBS \times CXT \times \Delta$ to subsets of $COMPONENTS$, denoting CBR retrieval,
- *pl*, a plausibility measure for cases produced by *cbr*.

Definition 1:

A set A of abducible terms is an *abductive explanation* or an *abductive diagnosis* for OBS given $MODEL$ in the context CXT if and only if:

1. $A \cap rok(OBS) = \{\}$,
2. $MODEL \cup CXT \cup A \vdash m$, for all $m \in \psi^+$,
3. $MODEL \cup CXT \cup A \cup \psi^-$ is consistent.

A is a *minimal abductive explanation* or a *minimal abductive diagnosis* if and only if there exists no proper subset of A that is an abductive explanation.

This definition of abductive explanation corresponds to the one of Console *et al.*'s model except for the use of the rule-out knowledge. The additional constraint (1) enforces that no component given as a potential diagnosis has been ruled-out by *rok*.

Example 1:

E_1 = { oil_cup_holed } in Figure 1 is an abductive diagnosis for both Session 1 and Session 2.

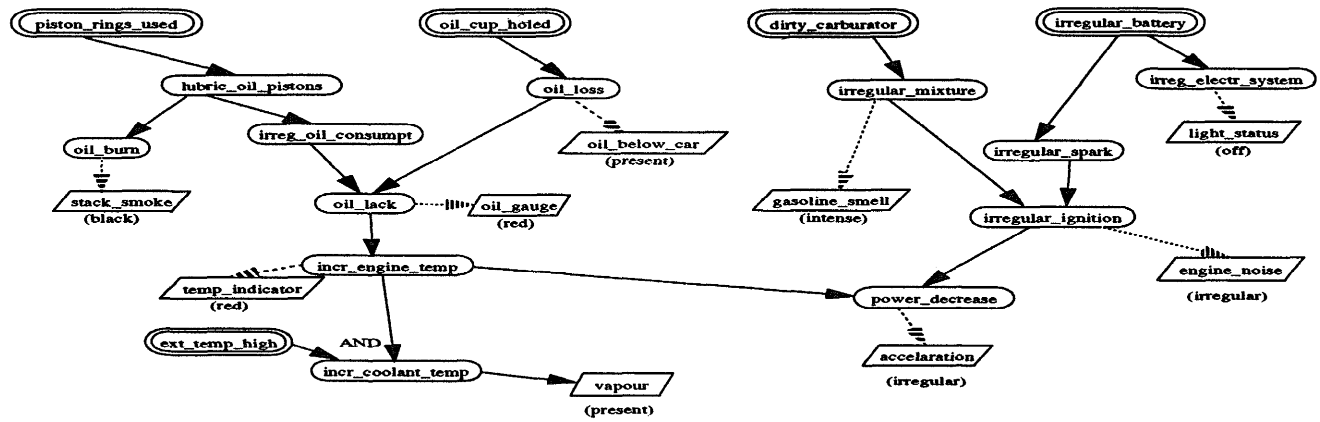
Definition 2:

A set $I \in cbr(OBS, CXT, \Delta)$ of abducible terms, is an *inductive explanation* for OBS given $MODEL$ in the context CXT if and only if there exists sets of ground terms δ_i^+ and δ_i^- such that:

1. $\delta_i^+ \subseteq \psi^+$ and $\delta_i^- \subseteq \psi^-$,
2. $MODEL \cup CXT \cup I \vdash_{pl} m$, for all $m \in \delta_i^+$,
3. $MODEL \cup CXT \cup I \cup \delta_i^-$ is consistent.

Inductive explanations depend on the current diagnostic situation (represented by OBS and CXT) and on the cases already in the case base Δ. An inductive explanation is weaker than an abductive explanation: it is only believed to entail some part of the observations $\delta_i^+ \subseteq \psi^+ \subseteq OBS$ that abductive explanation must entail, and is required to be consistent with less of the data in ψ^- than abductive explanations ($\delta_i^- \subseteq \psi^-$).

The sets ψ^+ and ψ^- depend on a specific application, while the sets δ_i^+ and δ_i^- depend on each case in Δ. Depending on the content of the case base, two cases can be exhibited with different sets δ_i^+ and δ_i^-. We leave the plausibility measure unspecified for now. This measure should depend, to some degree, on the size of the sets δ_i^+ and δ_i^-, since the more a hypothesis can explain, the more it is likely to be useful for the final diagnosis. Note that the definition of inductive explanations makes the inductive process of EAD weaker than the abductive phase. This corresponds to the intuitive notion that nothing guarantees the relevance of a past case to the current situation, whereas the device model and a sound abductive process produce sound explanations for the symptoms at hand.



Session 1:

$OBS = \{$ vapour(present), acceleration(irregular), stack_smoke(normal), gasoline_smell(normal), light_status(on) $\}$

$\psi^+ =$ { vapour(present), acceleration(irregular) }

$\psi^- =$ {¬vapour(absent), ¬acceleration(regular), ¬stack_smoke(black), ¬gasoline_smell(intense), ¬light_status(off) }

Session 2:

$OBS = \{$ vapour(present), stack_smoke(normal), gasoline_smell(normal), light_status(on) $\}$

$\psi^+ =$ { vapour(present)}

$\psi^- =$ {¬vapour(absent), ¬stack_smoke(black), ¬gasoline_smell(intense),¬light_status(off) }

Figure 1: Console *et al.*'s model and two diagnostic sessions

Definition 3:

An inductive explanation I is an *inductive diagnosis* if the plausibility measure for I is greater than or equal to a prescribed threshold P.

Definition 4:

A *case base* Δ is defined as a set of past diagnostic cases: $\Delta = \{\Delta_i = (OBS_i, A_i, I_i, \varepsilon_i)\}$, where OBS_i is a set of observations, A_i is the set of abductive diagnoses for OBS_i, I_i is the set of inductive diagnoses for OBS_i, and ε_i is the correct diagnosis for OBS_i, such that $\varepsilon_i \not\subseteq A_i$.

Because the correct diagnosis ϵ_i is not produced by the abductive part of the system ($\varepsilon_i \not\subseteq A_i$), Δ only contains cases for which the abductive explanation process failed, i.e. failure cases.

Example 2:

If both Sessions 1 and 2 failed, the case base Δ would contain two cases, representing both sets of observations, the sets of abductive and the inductive diagnoses produced for them, and the correct diagnosis E_1 for both cases[1].

The model for EAD covers both consistency-based and abduction-based explanations, and is complex to use. For the sake of simplicity, we will use a simpler functional notation: $Model(A)$ denotes the deductive closure of the conjunct $MODEL \cap CXT \cap A$, and $Model^{-1}(OBS)$ represents the set of abductive diagnoses found for a set of observations OBS. Similarly, $\mathcal{M}$ denotes the perfect device model that is always consistent and complete with respect to reality. If A is a set of ground terms denoting some facts, $\mathcal{M}(A)$ is the deductive closure of A through $\mathcal{M}$, i.e. all true facts that can be correctly deduced from A using $\mathcal{M}$. $\mathcal{A}_{pos}$ denotes the set of individual hypotheses that can truly explain the observations OBS. For all possible OBS, the following relationship holds: $OBS \subseteq \mathcal{M}(\mathcal{A}_{pos})$, asserting that, given the correct diagnoses, $\mathcal{M}$ correctly explains at least the observations OBS.

The inverse $\mathcal{M}^{-1}$ of $\mathcal{M}$ represents the ultimate abductive diagnoser, the one that the implemented system attempts to approximate. The complete diagnostic truth, i.e. all possible correct diagnoses for a observations OBS, is represented by $\mathcal{M}^{-1}(OBS)$. If ε_i represents the final correct diagnosis for a set of observations OBS_i, then the following relation holds: $\varepsilon_i \subseteq \mathcal{M}^{-1}(OBS) = \mathcal{A}_{pos}$.

We are interested in studying how the CBR system improves the overall performance of the diagnostic system. Thus, it is necessary to characterize diagnostic errors first and study how the CBR system can account for these errors. We next introduce additional definitions concerning the abductive part of the model for EAD.

Definition 5:

[1] We are only concerned with diagnostic failures here. Successful cases could also be used effectively especially to improve the ranking on the potential diagnoses produced by the abductive part of the diagnoser.

rok is *covering* if
$$\mathcal{A}_{pos} \subseteq (COMPONENTS - rok(OBS)).$$
rok is covering if it does not forget hypotheses, i.e. if nothing that can explain the set of observations OBS is pruned away by *rok*.

Example 3:

A set of rules containing the rule:

IF engine_noise(irregular) THEN ¬irregular_battery

would not be covering, since it would discard the possibility that the battery might be faulty, prematurely concentrating the abductive process on the carburator.

Definition 6:

$Model$ is *faithful* if,
$$\forall OBS, \forall A \in Model^{-1}(OBS),$$
$$Model(A) \subseteq \mathcal{M}(A).$$
$Model$ is faithful if, for all abductive diagnoses A for OBS, all data entailed by $Model(A)$ is correct with respect to $\mathcal{M}$.

Example 4:

This example assumes the model in Figure 1 except that the two nodes *oil_burn* and *oil_loss* are switched along with their associated symptoms *oil_below_car* and *stack_smoke*(*black*). It also assumes that this model is correct i.e. corresponds to $\mathcal{M}$. Given the observations from Session 1, the diagnoser would output *piston_rings_used* as the only possible abductive diagnosis. In turn, we have
$$Model(piston\_rings\_used) \not\subseteq \mathcal{M}(piston\_rings\_used),$$
showing that $Model$ is not faithful.

Definition 7:

$Model$ is *complete* if,
$$\forall OBS, \forall A \in Model^{-1}(OBS), \mathcal{M}(A) \subseteq Model(A).$$
$Model$ is complete if it explains all the data in OBS that is explainable (with respect to $\mathcal{M}$).

Example 5:

Example 4 above is an example where $Model$ is not complete since
$$piston\_rings\_used \in \mathcal{M}^{-1}(OBS) \text{ and},$$
$$\mathcal{M}(piston\_rings\_used) \not\subseteq Model(piston\_rings\_used).$$

Different ways of using inductive diagnoses can be designed depending on the characteristics of each application. For example, a rough estimate of the quality of the model could be used to determine the degree with which to use and trust the inductive diagnostic process. The more the device model can be trusted, the less the past cases are needed for the current diagnostic situation. Another consideration is the similarity measure used to compare the current situation with retrieved past cases. A statistical analysis of this measure could greatly help determine the uncertainty of its results. Any procedure for merging abductive and inductive diagnoses would be somewhat device dependent and does not belong to a general model for diagnosis combining abduction and induction.

Characterizing Diagnostic Errors

Given the model for diagnosis defined above, and the definitions in the previous section, an error-free diagnoser is one such that all possible explanations for the observations OBS are produced, and that all produced explanations are correct, i.e.:
$$OBS \subseteq \mathcal{M}(\mathcal{A}_{pos}) \subseteq Model(rok(OBS)) \subseteq \mathcal{M}(rok(OBS)) \quad (1)$$
This implies three types of possible diagnostic errors:

Type A: the system fails to explain the data OBS: $OBS \not\subseteq Model(rok(OBS))$

Type B: the system explains the data but in an incorrect way: $Model(rok(OBS)) \not\subseteq \mathcal{M}(rok(OBS))$,

Type C: possible explanations are overlooked: $\mathcal{M}(\mathcal{A}_{pos}) \not\subseteq Model(rok(OBS))$.

Theorem 1:

If $Model$ is monotonic, *rok* is covering, $Model$ is complete, and $Model$ is faithful then there will be no diagnostic errors.

Equation 1 imposes that for all possible OBS, the data in OBS be explained according to the definition of diagnosis given in the previous section, that all explanations be correct, and that all possible explanations be produced. $Model$ might have to explain more data than there is in OBS. These "extraneous" explanations, which do not explain OBS directly, still have to be correct for diagnostic errors to be avoided. Equation 1 constrains the restriction of $Model$ to all possible $rok(OBS)$, which might be different from $COMPONENTS$. This is important in the case of cancellation effects between hypotheses, or when two pieces of data cannot be present separately.

If $\mathcal{M}$ is not monotonic, i.e. if $\exists A, \exists A'$, $A' \subseteq A$ and $\mathcal{M}(A') \not\subseteq \mathcal{M}(A)$ more hypotheses can potentially explain less symptoms or observations. Intuitively, this situation is more difficult than if $\mathcal{M}$ is monotonic since it implies that consequences of errors in the device model either create erroneous explanations (as if $Model$ was monotonic) or forget explanations and diagnoses (for example if there is an erroneous cancellation effect). A non-monotonic abduction problem makes the search for minimal hitting sets for minimal conflicts NP-hard (Bylander *et al.* 1991).

We have showed the conditions that guarantee the absence of diagnostic errors. Figure 2 shows what kind of errors are produced if these conditions are not met. It describes eight situations combining the three types of errors described above. Errors of type B can occur if $Model$ is not faithful, i.e. incorrect explanations can only be produced by an incorrect model. If the whole device model is considered, i.e. if *rok* is covering, then errors of type A and C can occur if $Model$ is not complete (Situations 3 and 4). Proofs of these results can be found in (Féret 1993). These results show that,

rok	Model		no	error types
covering	complete	faithful	1	
		not faithful	2	B
	not complete	faithful	3	A C
		not faithful	4	A B C
not covering	complete	faithful	5	A C
		not faithful	6	A B C
	not complete	faithful	7	A C
		not faithful	8	A B C

Figure 2: Summary of diagnostic error situations

in general, it is hard to discriminate between potential causes of diagnostic errors, directly from the types of errors that are produced. However, specific conditions might render performance improvement easier. For example, if we assume that *rok* is covering, then the problem seems simpler: if an error of type B occurs, then *Model* is not faithful, and if an error of type A or C occurs, *Model* is at least not complete. This gives a strategy to try to improve the device model.

The results above show that there is little hope for a general abductive method for trouble-shooting diagnostic systems. This is a strong argument in favor of alternative methods which aim at improving the performance of MBD systems. This retrospectively justifies the EAD approach.

Discussion

The definitions given in the paper were inspired by the work on the complexity of abduction done by Bylander *et al.* (Bylander *et al.* 1991) and relate to the focused abductive diagnosis of Console *et al.* (Console, Portinale, & Theseider Dupre 1991) and to Reiter's theory of diagnosis from first principles (Reiter 1987).

The mapping from the EAD model to Console *et al.*'s model is as follows:

$$OBS = D,$$
$$rok(OBS) = C \subseteq COMPONENTS,$$
$$cbr(OBS, CXT, \Delta) = \{\},$$

Similar to EAD, their model does not make assumptions about the device models it uses. The EAD model is also interesting because it uses compiled knowledge to rule-out possible hypotheses (explanations). This compiled knowledge takes the form of necessary conditions (for parts of the model to be considered) and are compiled *a priori* from the behavioral model of the device. The EAD approach also addresses the control problem of choosing between abductively generating more hypotheses and deductively eliminating others that is present in Console *et al.*'s model. The rule-out knowledge is applied first and the remaining hypotheses are worth investigating and represent a lesser waste of effort should they be found impossible. The EAD approach is therefore conceptually simpler. The importance of rule-out knowledge is illustrated by considering its role in pruning away individual hypotheses. A linear decrease in the number n of possible individual hypotheses corresponds to an exponential decrease of the size 2^n of the superset of the set of possible individual hypotheses, i.e. the size of the search space. The example in Figure 1 illustrates the potential usefulness of the rule-out knowledge: if the fact $engine\_noise(regular)$ is known, then the following can be logically inferred: 1) the battery is fine, 2) the carburator is not dirty, 3) the ignition is regular, and 4) if there is a power decrease, it can only be because the engine is too hot. This is effective rule-out knowledge since one fact rules out a whole branch of the search space, therefore significantly reducing the cost of the abductive phase of EAD.

The EAD model does not make assumptions on how the rule-out knowledge (*rok*) is brought into the system. It can be compiled from the a deep model (as in (Console, Portinale, & Theseider Dupre 1991)) or simply hand coded by human experts (as in (Milne 1987)). One difference between our model and that of Console *et al.* is that we use rule-out knowledge *before* the abductive process. Koton's CASEY also uses CBR before a more traditional diagnostic approach to achieve speed-up learning - a simpler learning than the one achieved by the EAD model (Koton 1988). Our approach improves efficiency by using rule-out knowledge as soon as possible. We believe that this is effective for many real-world devices, because it is easy to formulate knowledge about independence (as mentioned in (Pearl 1986)) and about inconsistencies between symptoms and potential causes for those symptoms.

Reiter's model is subsumed by Console *al.*'s but provides insights relative to the search for minimal conflicts and minimal hitting sets. The mapping from our model to Reiter's is as follows:

$$OBS = D,$$
$$rok(OBS) = D,$$
$$cbr(OBS, CXT, \Delta) = \{\},$$
$$\psi^+ = \{\}.$$

For Reiter, a solution to the diagnosis problem is to find a diagnosis for OBS which satisfies the consistency requirement. There is no equivalent of the plausibility function in Reiter's model, i.e. no way of ranking diagnoses.

In the EAD model, the rule-out knowledge *rok* can be seen as a way to restrict the search for minimal conflicts. If knowledge about impossibilities of some components to be faulty, or of some combinations of components to be faulty simultaneously is available, then the search for minimal conflicts (minimal sets of components that cannot be simultaneously in a normal state together, given the observations OBS) will be reduced. However, the search for minimal hitting sets for the minimal conflicts might actually be made more complex by cancellation knowledge as defined by

Bylander *et al.*, making this search NP-hard. The deductive, rule-out phase of the EAD model addresses this problem by forcing the system to discard as many impossible faults as possible before considering combinations of possible ones. This phase could typically use compiled knowledge to reduce the expensive use of deep knowledge.

Abductive induction, implemented by a CBR system, is used to recall past errors and to avoid these errors in similar situations. Assuming that each failure case is stored in Δ, and that the information derived from the abductive diagnostic process serves as indices for the cases, inductive diagnoses can be retrieved, when the current situation is similar enough to cases stored in Δ, to correct errors made in previous situations. The indices used by the CBR system are paths (or portions of paths) followed by the abductive process through the device model to derive diagnoses. This insures the relevance to the current situation of cases stored under these paths. The building of the case base Δ can be seen as an *a fortiori* knowledge compilation process. Because this process takes place when the system is in place, it produces more focused and more useful compiled knowledge than knowledge compiled *a priori* (Console, Portinale, & Theseider Dupre 1991). (Féret & Glasgow 1993; Féret 1993) presented experimental results that support the effectiveness of the EAD model, especially of its inductive phase.

The EAD model has been implemented as part of the Automated Data Management System (ADMS) and has been applied to two real-world devices: a robotic system called the Fairing Servicing Subsystem, and a Reactor Building Ventilation System (Féret & Glasgow 1991).

This paper makes explicit the relation between CBR and induction by formally defining the notion of abductive induction in the context of automated diagnosis. It presented the EAD model which addresses the problems of computational complexity and of incorrect device models. The EAD model is a practical and general answer to problems encountered while trying to apply MBD to real-world situations. The research presented in this paper also contributes to the field of CBR by providing a formal model of a hybrid CBR system which characterizes the interactions of CBR with another problem-solving paradigm.

References

Bylander, T.; Allemang, D.; Tanner, M. C.; and Josephson, J. R. 1991. The computational complexity of abduction. *Artificial Intelligence* 49:25–60.

Console, L.; Portinale, L.; and Theseider Dupre, D. 1991. Focusing abductive diagnosis. *AI Communications* 4(2/3):88–97.

Console, L.; Theseider Dupre, D.; and Torasso, P. 1989. A theory of diagnosis for incomplete causal models. In *Proceedings of 11th IJCAI*, 1311–1317.

de Kleer, J., and Williams, B. C. 1987. Diagnosis multiple faults. *Artificial Intelligence* 32:97–129.

de Kleer, J. 1989. Diagnosis with behavioral modes. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, 1324–1330.

de Kleer, J. 1990. Using crude probability estimates to guide diagnosis. *Artificial Intelligence* 45(3):381–391.

de Kleer, J. 1992. Optimizing focusing model-based diagnosis. In *Proceedings of the 3rd International Workshop on Principles of Diagnosis*, 26–29.

Féret, M. P., and Glasgow, J. I. 1991. Generic diagnosis for mechanical devices. In *Proceedings of the 6th International Conference on Applications of Artificial Intelligence in Engineering*, 753–768. Oxford, UK: Computational Mechanics Publications, Elsevier Applied Science.

Féret, M. P., and Glasgow, J. I. 1993. Hybrid case-based reasoning for the diagnosis of complex devices. In *Proceedings of AAAI-93*, 168–175.

Féret, M. P. 1993. *Explanation-Aided Diagnosis: Combining Case-Based and Model-Based Reasoning for the Diagnosis of Complex Devices.* Ph.D. Dissertation, Department of Computing and Information Science, Queen's University, Kingston, Ontario, Canada.

Friedrich, G. 1992. Theory diagnoses: A concise characterization of faulty systems. In *Proceedings of the 3rd International Workshop on Principles of Diagnosis*, 117–131.

Koton, P. 1988. Reasoning about evidence in causal explanations. In *Proceedings of AAAI-88*, 256–261.

Milne, R. 1987. Strategies for diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-17(3):333–339.

Mozetic, I. 1990. Reduction of diagnostic complexity through model abstractions. In *Proceedings of the 1rst International Workshop on Principles of Diagnosis*, 102–111.

Pearl, J. 1986. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence* 28:241–288.

Peirce, C. S., ed. 1955. *Philosophical Writings of Peirce.* New York: Dover Publications Inc.

Reiter, R. 1987. A theory of diagnosis from first principle. In *Artificial Intelligence*, volume 32, 57–95.

Riesbeck, C., and Schank, R., eds. 1989. *Inside Case-Based Reasoning.* Lawrence Erlbaum Associates.