A Model Counting Characterization of Diagnoses

T. K. Satish Kumar

Knowledge Systems Laboratory
Stanford University
tksk@ksl.stanford.edu

Abstract

Given the description of a physical system in one of several forms (a set of constraints, Bayesian network etc.) and a set of observations made, the task of model-based diagnosis is to find a suitable assignment to the modes of behavior of individual components (this notion can also be extended to handle transitions and dynamic systems [Kurien and Nayak, 2000]. Many formalisms have been proposed in the past to characterize diagnoses and systems. These include consistency-based diagnosis, fault models, abduction, combinatorial optimization, Bayesian model selection etc. Different approaches are apparently well suited for different applications and representational forms in which the system description is available. In this paper, we provide a unifying theme behind all these approaches based on the notion of model counting. By doing this, we are able to provide a universal characterization of diagnoses that is independent of the representational form of the system description. We also show how the shortcomings of previous approaches (mostly associated with their inability to reason about different elements of knowledge like probabilities and constraints) are removed in our framework. Finally, we report on the computational tractability of diagnosis-algorithms based on model counting.

1 Introduction

Diagnosis is an important component of autonomy for any intelligent agent. Often, an intelligent agent plans a set of actions to achieve certain goals and because some conditions may be unforeseen, it is important for it to be able to reconfigure its plan depending upon the state in which it is. This state identification problem is essentially a problem of diagnosis. In its simplest form, the problem of diagnosis is to find a suitable assignment to the modes of behavior of individual components in a static system (given some observations made on it). It is possible to handle the case of dynamic systems by treating the transition variables as components (in one sense) [Kurien and Nayak, 2000]. The theory developed in this paper is therefore equally applicable to dynamic systems too

(although we omit the discussion due to restrictions on the length of the paper).

Many approaches have been used in the past to characterize diagnoses and systems. Among the most comprehensive pieces of work are [de Kleer and Williams, 1989], [Reiter, 1987], [Struss and Dressler, 1989], [Console *et al.*, 1989], [de Kleer *et al.*, 1992], [Poole, 1994], [Kohlas *et al.*, 1998] and [Lucas, 2001]. The popular characterizations of diagnoses include consistency-based diagnosis, fault models, abduction, combinatorial optimization, and Bayesian model selection. These approaches are however tailored for different applications and representational forms in which the system description is available. They also have one or more shortcomings arising out of their inability to provide for a framework that can incorporate knowledge in different forms like probabilities, constraints etc.

In this paper, we provide a unifying theme behind all these approaches based on the notion of model counting. By doing this, we are able to provide a universal characterization of diagnoses independent of the representational form of the system description. Because model counting bridges the gap between different kinds of knowledge elements, the shortcomings of previous approaches are removed.

2 Background

Before we present our characterization of diagnoses based on model counting, we choose to provide a quick overview of the previous approaches so that we can compare and contrast our approach with them.

Definition *(Diagnosis System)* A *diagnosis system* is a triple $(SD, COMPS, OBS)$ such that:

1. SD is a system description expressed in one of several forms — constraint languages like propositional logic, probabilistic models like Bayesian network etc. SD specifies both component behavior information and component structure information (i.e. the topology of the system).

2. $COMPS$ is a finite set of components of the system. A component $comp_i$ ($1 \leq i \leq |COMPS|$) can behave in one of several, but finite set of modes (M_i). If these modes are not specified explicitly, then we assume two modes — failed ($AB(comp_i)$) and normal ($\neg AB(comp_i)$).

3. OBS is a set of observations expressed as variable values.

Definition The task in a *complete diagnosis call* is to find a "suitable" assignment of modes to all the components in the

system given SD and OBS. The task in a *partial diagnosis* call is to find a suitable assignment of modes to a specified subset S ($S \subseteq COMPS$) of the components in the system given SD and OBS.

Unless stated otherwise, we will use the term "diagnosis" to refer to a complete diagnosis. Later in the paper we will show that the characterization of partial diagnoses is a simple extension of the characterization of complete diagnoses.

Definition *(Candidate)* Given a set of integers $i_1 \cdots i_{|COMPS|}$ (such that for $1 \leq j \leq |COMPS|$, $1 \leq i_j \leq |M_j|$), a candidate $Cand(i_1 \cdots i_{|COMPS|})$ is defined as $Cand(i_1 \cdots i_{|COMPS|}) = (\bigcup_{k=1}^{|COMPS|}(comp_k = M_k(i_k)))$.
Here, $M_u(v)$ denotes the vth element in the set M_u (assumed to be indexed in some way).

Notation When the indices are implicit or arbitrary, we will use the symbol H to denote a candidate or a hypothesis i.e. an assignment of modes to all the components in the system.

Consistency-Based Diagnosis

The task of consistency-based diagnosis can be summarized as follows. Note that the definition of a diagnosis in this framework does not discriminate between single and multiple faults.

Definition *(Consistency-Based Diagnosis)* A candidate H is a diagnosis if and only if $SD \cup OBS \cup H$ is satisfiable.

There are other characterizations of diagnoses under this framework called *partial diagnoses*, *prime diagnoses*, *kernel diagnoses* etc. We will examine these later in the paper.

Fault Models

Consider diagnosing a system consisting of three bulbs B_1, B_2 and B_3 connected in parallel to the same voltage source V under the observations $off(B_1)$, $off(B_2)$ and $on(B_3)$. $AB(V) \wedge AB(B_3)$ is a diagnosis under the consistency-based formalization of diagnosis if we had constraints only of the form $\neg AB(B_3) \wedge \neg AB(V) \to on(B_3)$. Intuitively however, it does not seem reasonable because B_3 cannot be *on* without V working properly. One way to get around this is to include fault models in the system. These are constraints that explicitly describe the behavior of a component when it is not in its nominal mode (most expected mode of behavior of a component). Such a constraint in this example would be $AB(B_3) \to off(B_3)$. Diagnosis can become indiscriminate without fault models. It is also easy to see that the consistency-based approach can exploit fault models (when they are specified) to produce more intuitive diagnoses (like only B_1 and B_2 being abnormal).

Diagnosis as Combinatorial Optimization

The technique of using fault models is associated with the problem of being too restrictive. We may not be able to model the case of some strange source of power making B_3 *on* etc. The way out of this is to allow for many modes of behavior for each component of the system. Every component has a set of modes (in which it can behave) with associated models. One of these is the nominal (or normal) mode and the others are fault modes. Each component has the *unknown* fault mode with the empty model. The unknown mode tries to capture the *modeling incompleteness assumption* (obscure

modes that we cannot model in the system). Also, each mode has an associated probability that is the prior probability of the component being in that mode. Diagnosis can now be cast as a combinatorial optimization problem of assigning modes of behavior to each component such that it is not only consistent with $SD \cup OBS$, but also maximizes the product of the prior probabilities associated with those modes (assuming independence in the behavior of components).

Definition *(Combinatorial Optimization)* A candidate $H = Cand(i_1 \cdots i_{|COMPS|})$ is a diagnosis if and only if $SD \cup H \cup OBS$ is satisfiable and $P(H) = (\Pi_{k=1}^{|COMPS|}P(comp_k = M_k(i_k)))$ is maximized.

Diagnosis as Bayesian Model Selection

Sometimes we have sufficient experience and statistical information associated with the behavior of a system. In such cases, the system description is usually available in the form of a probabilistic model like a Bayesian network. Given some observations made on the system, the problem of diagnosis then becomes a Bayesian model selection problem.

Definition *(Bayesian Model Selection)* A candidate H is a diagnosis (for a probabilistic model of the system, SD) if and only if it maximizes the posterior probability $P(H/SD, OBS)$.

Diagnosis as Abduction

Yet another intuition behind characterizing diagnoses is the idea of explanation. Explanatory diagnoses essentially try to capture the notion of cause and effect in the physics of the system. The observations are asymmetrically divided into inputs (I) and outputs (O) [de Kleer *et al.*, 1992]. The inputs (I) are those observation variables that can be controlled externally.

Definition *(Abductive Diagnosis)*: An abductive diagnosis for $(SD, COMPS, OBS = I \cup O)$ is a candidate H such that $SD \cup I \cup H$ is satisfiable and $SD \cup I \cup H \to O$.

3 Probabilities and Model Counting

Before we present our own characterization of diagnoses based on the notion of model counting, we show an interesting relationship between probabilities and model counting (see Figure 1). The model counting problem is the problem of counting the number of solutions to a SAT (satisfiability problem) or a CSP (constraint satisfaction problem).

Definition *(Binary representation of a CPT)*: The *binary representation of a CPT (Conditional Probability Table)* is a table in which all the floating-point entries of the CPT are re-written in a binary form (base 2) up to a precision of P binary digits and the decimal point along with any redundant zeroes to the left of it are removed.

We provide a set of definitions and results relating the probability of a partial assignment A to the number of solutions (under the same partial assignment A) to CSPs composed out of the binary representations of the CPTs (see Figure 1). Basic definitions related to CSPs can be found in [Dechter, 1992].

Definition *(Zero-one-layer of a CPT)* The kth *zero-one-layer of a CPT* is a table of zeroes and ones derived from the kth
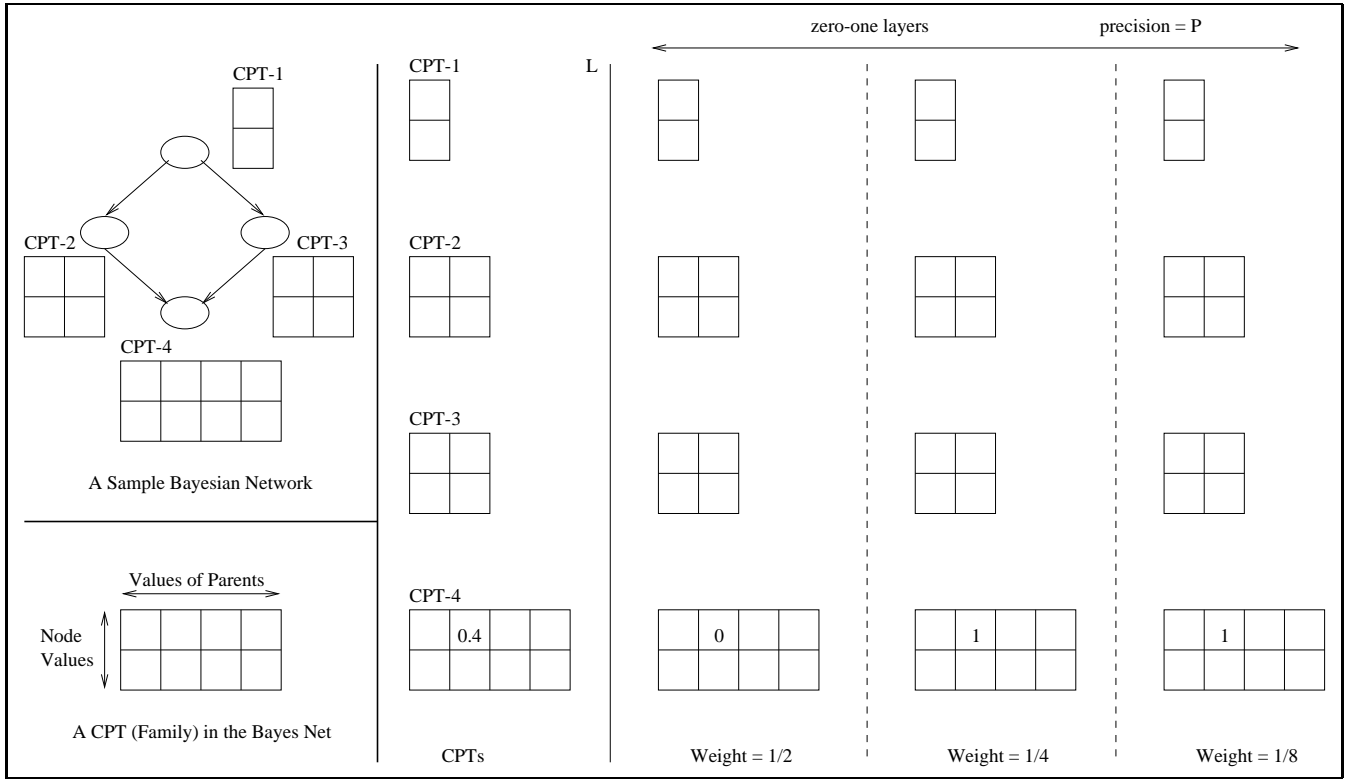
Figure 1: Shows the conditional probability tables (CPTs) of a Bayes net on the left of the vertical line L. On the right of L are the binary representations of these CPTs (example shown for 0.4 in decimal = 0.011 in binary). CPTs correspond to families in the Bayes net and let the number of families be C.

bit position of all the numbers in the binary representation of that CPT.

Definition *(Weight of a zero-one-layer)* The kth zero-one-layer of a CPT is defined to have weight 2^{-k}.

Definition *(CSP Compilation of a CPT)* The kth *CSP compilation of a CPT* is a constraint over the variables of the CPT that is derived from the kth zero-one-layer of the CPT such that zeroes correspond to disallowed tuples and ones correspond to allowed tuples.

Definition *(CSP Compilation of Network)* The $(k_1, k_2 \cdots k_C)$ *CSP compilation of the Network* is the set of constraints $S = \{s_i : s_i$ is the k_i^{th} CSP compilation of the ith CPT$\}$.

Definition *(Weight of a CSP Compilation)* The *weight of a* $(k_1, k_2 \cdots k_C)$ *CSP compilation* of a network is defined to be equal to $2^{-(k_1+k_2\cdots k_C)}$.

Property There are an exponential number of CSP compilations for a given network. Since each CPT expands into P zero-one-layers and a CSP for the entire network can be compiled by taking any of these P layers for each CPT (there are C CPTs), the total number of CSP compilations possible is P^C.

Notation We will use the notation h_{ij} to mean the jth CSP compilation of the ith CPT. Let A indicate a complete or partial assignment to the variables. If A is an assignment that instantiates all the variables of CPT_i, then we will use the notation $h_{ij}(A)$ to indicate whether or not A satisfies h_{ij}. If A is a complete assignment for all the variables in

the network, then all variables for all CPTs are instantiated and we will use the notation $CSP_{(k_1,k_2\cdots k_C)}(A)$ to indicate whether A satisfies all the constraints h_{ik_i} ($1 \leq i \leq C$). If A is not a complete assignment for all the variables, then we will use the notation $\#CSP_{(k_1,k_2\cdots k_C)}(A)$ to indicate the number of solutions to the $(k_1, k_2 \cdots k_C)$ CSP compilation of the network that share the same partial assignment as A.

Theorem 1 The probability of a complete assignment $A = (X_1 = x_1 \cdots X_n = x_n)$ is just the sum of the weights of the different CSP compilations of the network that are satisfied by this complete assignment. That is, $P(A) = \sum_{(k_1,k_2\cdots k_C)} CSP_{(k_1,k_2\cdots k_C)}(A)2^{-(k_1+k_2\cdots k_C)}$ (for all $1 \leq i \leq C, 1 \leq k_i \leq P$).

Proof Consider the complete assignment $A = (X_1 = x_1 \cdots X_n = x_n)$ for all the variables. The probability of this assignment is equal to the product of the probabilities defined *locally* by each CPT. Now using the fact that the tth bit in the binary representation of this local value has been written out as an allowed or disallowed tuple in the tth CSP compilation of that CPT, we can rewrite the local value for A in a CPT as $\sum_{j=1}^{P} h_{ij}(A)2^{-j}$. The total probability is then just the product over all local values $= \Pi_{k=1}^{C} \sum_{j=1}^{P} h_{kj}(A)2^{-j}$. Expanding the product, we see that each term is essentially of the form $\sum_{(k_1,k_2\cdots k_C)} 2^{-(k_1+k_2\cdots k_C)} \Pi_{j=1}^{C} h_{jk_j}(A) = \sum_{(k_1,k_2\cdots k_C)} 2^{-(k_1+k_2\cdots k_C)} CSP_{(k_1,k_2\cdots k_C)}(A)$.

Theorem 2 *(Model Counting)* The marginalized probability of a partial assignment A to a set of variables $S \subseteq V$ is equal to the product of the weight and the number of solutions (under the same partial assignment A) summed over all CSP compilations of the network. That is, $P(A) = \sum_{(k_1,k_2\cdots k_C)} \#CSP_{(k_1,k_2\cdots k_C)}(A) 2^{-(k_1+k_2\cdots k_C)}$ (for all $1 \leq i \leq C, 1 \leq k_i \leq P$).

Proof From the previous theorem, we know that the probability of a complete assignment B is $\sum_{(k_1,k_2\cdots k_C)} CSP_{(k_1,k_2\cdots k_C)}(B) 2^{-(k_1+k_2\cdots k_C)}$ (for all $1 \leq i \leq C, 1 \leq k_i \leq P$). Now, the marginalized probability of a partial assignment A is just the sum of the probabilities of all complete assignments B that agree with A on the assignment to variables in S. That is, $P(A) = \sum_B P(B)(B(S) = A)$. Using the result of the previous theorem to expand $P(B)$, we have $P(A) = \sum_B \sum_{(k_1,k_2\cdots k_C)} CSP_{(k_1,k_2\cdots k_C)}(B) 2^{-(k_1+k_2\cdots k_C)}(B(S) = A)$. Switching the two summations and noting that $\sum_B CSP_{(k_1,k_2\cdots k_C)}(B)(B(S) = A)$ is the same as $\sum_{(k_1,k_2\cdots k_C)} \#CSP_{(k_1,k_2\cdots k_C)}(A)$, we get that $P(A) = \sum_{(k_1,k_2\cdots k_C)} \#CSP_{(k_1,k_2\cdots k_C)}(A) 2^{-(k_1+k_2\cdots k_C)}$.

3.1 Probability-Equivalents and Incorporation of Probabilities

Often, we are given information in many forms. Probabilities are natural information elements when there is an element of statistical experience that we want to exploit. In other cases, constraints may be the most appropriate to use. The general idea in our framework is to use probabilities when we explicitly have them and to use model counting otherwise. We will use $\#(S_1, S_2 \cdots)$ to mean the number of consistent models to $(S_1 \cup S_2 \cdots)$ (with respect to the uninstantiated free variables in SD). Theorems 1 and 2 establish that model counting is a weaker form of probabilities and that probabilities provide only *precision information* over model counting. Therefore, it is natural for us to use probabilities (to describe events) when we have them explicitly, and to use model counting otherwise. For any event E, we use the expressions $\frac{\#(SD,E)}{\#(SD)}$ and $P(E)$ almost equivalently — except that we use the former when we do not know $P(E)$ explicitly. This framework allows us to reason about both probabilities and constraints.

Definition *(Probability Equivalents)* The *probability equivalent* of $\#(SD, E)$ for any event E is defined to be $P(E)\#(SD)$ when $P(E)$ is given explicitly.

4 Diagnosis as Model Counting

In this section, we characterize diagnoses based on model counting. We will then show how all the previous approaches are captured under this formalization. For the first part of the discussion we will consider only complete diagnoses (an assignment of modes for all the components).

Definition *(Model Counting Characterization)* A *diagnosis* is a candidate H that maximizes the number of consistent models to $SD \cup OBS \cup H$ using probability equivalents wherever necessary.

Notation We will use $M(H)$ to denote $\#(SD, OBS, H)$ (the number of consistent models to $SD \cup OBS \cup H$) when

SD and OBS follow from context.

Theorem 3 *(Capturing Consistency-Based Diagnosis)* Consistency-Based diagnosis is looking for a hypothesis H for which $M(H)$ is non-zero.

Proof By definition, consistency-based diagnosis chooses H such that $SD \cup OBS \cup H$ is consistent. In other words, there exists at least one satisfying assignment for $SD \cup OBS \cup H$. Clearly, this is equivalent to saying that $M(H)$ is non-zero.

Theorem 4 *(Capturing Abduction)* Abduction chooses a hypothesis H that maximizes $M(H)$ assuming uniformity in prior probabilities $P(H)$.

Proof The maximum value of $\#(SD, OBS = I \cup O, H)$ is $\#(SD, H, I)$ and this happens when $H \cup SD \cup I \rightarrow O$. Given that the input variables are controlled externally, we know that $\#(SD, H) = N(I)\#(SD, H, I)$. Here, $N(I)$ is a constant that measures the number of different values for the input variables. Since $\#(SD, H)$ is equivalent to $P(H)\#(SD)$ which we assumed to be a constant for all H, maximizing $\#(SD, OBS, H)$ is equivalent to finding a hypothesis H for which $I \rightarrow O$ (under SD). The fact that abduction requires H to be consistent is also captured, because if H is inconsistent, then $M(H) = 0$ and clearly $M(H)$ will not be maximized.

Theorem 5 *(Capturing Bayesian Model Selection)* Bayesian model selection chooses a hypothesis H such that it maximizes the probability equivalent of $M(H)$.

Proof The probability equivalent of $M(H) = \#(SD, OBS, H)$ is $P(OBS, H)$. Clearly, if we are maximizing $P(OBS, H)$ then we are maximizing $P(H/OBS)P(OBS)$. Since $P(OBS)$ is independent of H, it is equivalent to maximizing $P(H/OBS)$ which is exactly what Bayesian model selection does.

Theorem 6 *(Capturing Combinatorial Optimization)* Combinatorial optimization is looking for a hypothesis H which maximizes $P(H)$ under the condition that $M(H)$ is non-zero.

Proof As noted earlier, H is consistent with $SD \cup OBS$ if and only if $M(H)$ is non-zero. We also know that combinatorial optimization is looking for a consistent H which maximizes $P(H)$. The theorem follows as a simple consequence of the above two statements. Basically, combinatorial optimization maximizes only the prior probabilities of hypotheses (instead of maximizing the equivalent of the posterior probabilities) unless they are obviously ruled out by being inconsistent.

4.1 Consequences (Removing Previous Shortcomings)

We now show the consequences of formalizing diagnosis as model counting. In particular, we identify problems with previous approaches and show how model counting removes all of them.

Problems with Consistency-Based Diagnosis

One of the problems with consistency-based diagnosis is that it allows for non-intuitive hypotheses as diagnoses. It provides only for a necessary but not a sufficient condition on the hypotheses that can be qualified as diagnoses. By itself, it is of little value unless we use an elaborate set of fault models

to remove non-intuitive hypotheses that could otherwise be consistent. Model counting removes these problems because of its ability to merge and incorporate the notions of both consistency and probabilities. In one sense, one can think of model counting as giving us a measure of the degree to which a hypothesis is consistent with SD and OBS. Some of these problems are alternatively addressed in [Kohlas *et al.*, 1998] and [Lucas, 2001].

Problems with Fault Models

The problem with fault-models is that of over-restriction (as explained at the beginning of the paper). We need to be able to reason not only about constraints relating SD and OBS, but also about any other kind of information we may have in the form of probabilities etc. The over-restriction problem can be removed by introducing probabilities. These probabilities can then be used in the unified framework of model counting.

Problems with Abduction

Like the consistency-based approaches, explanatory diagnoses are also unable to incorporate and reason about probabilities. Yet another problem with abduction is that it assumes we have completely modeled all cause-effect relationships in our system. This contradicts our modeling incompleteness assumption and is an unnecessary restriction on SD. Model counting removes this problem in a way very similar to how probabilities were used to deal with the modeling incompleteness assumption. Alternate treatments for these problems can be found in [Poole, 1994] (which links abduction with probabilistic reasoning) and [Console *et al.*, 1989] (which addresses the modeling incompleteness assumption).

Problems with Diagnosis as Bayesian Model Selection

Bayesian model selection agrees with our characterization of diagnoses — but the only problem it poses is that it requires SD to be in the form of a Bayesian network with known probabilities. Modeling a physical system as a Bayesian network is in many cases a non-intuitive thing to do. This is especially so when certain probability terms are hard to get. Parts of the system may be better expressed in the form of constraints or automata. In such cases, Bayesian model selection does not extend in a natural way and model counting is the right substitute (because it is defined under all frameworks).

Problems with Diagnosis as Combinatorial Optimization

One problem associated with casting diagnosis as a combinatorial optimization problem is that of being unable to give explanatory diagnoses a preference over the rest. Clearly, we would like to prefer hypotheses that not only maximize the prior probability $P(H)$ but that are also explanatory rather than just being consistent with $SD \cup OBS$. One way to incorporate this preference is to find all consistent hypotheses that maximize $P(H)$ and to pick an explanatory one among them. The question that arises then is how we would compare two hypotheses one of which is explanatory and the other just consistent (but not explanatory), with the latter having a slightly better prior probability. This question is left unanswered under the combinatorial optimization formulation of diagnoses. In the model counting framework however, it is easy to see

that we really have to maximize $P(H)\frac{\#(SD,OBS,H)}{\#(SD,H)}$. The second factor is maximized for explanatory diagnosis — but this is as much as the preference we attach for them.

Another problem with the combinatorial optimization formulation is that probabilities are restricted to only behavior modes of components and only these prior probabilities are maximized. There is no framework to reason about probabilistic information connected with observation variables.

5 Partial Diagnoses

Sometimes, we are interested in finding a suitable assignment of modes to a specified subset S of the components $COMPS$ rather than for all components. We argue that our characterization of diagnoses under the model counting framework remains unchanged.

Definition *(Candidate)* Given a set of integer tuples $(k_1, i_{k_1}) \cdots (k_n, i_{k_n})$ such that for $1 \leq j \leq n \leq |COMPS|$, $1 \leq i_{k_j} \leq |M_j|$, a candidate $Cand((k_1, i_{k_1}) \cdots (k_n, i_{k_n}))$ is defined as $Cand((k_1, i_{k_1}) \cdots (k_n, i_{k_n})) = (\bigcup_{g=1}^{n}(comp_g = M_g(i_{k_g})))$.

Notation When the indices are implicit or arbitrary, we will use the notation J_S to denote a candidate or a hypothesis i.e. a set of mode assignments to all the components in $S \subseteq COMPS$.

Definition *(Model Counting Characterization)* A *partial diagnosis* for $S \subseteq COMPS$ is an assignment of modes J_S to the components in S that maximizes $\#(SD, OBS, J_S)$ using probability equivalents wherever necessary.

It is now not hard to verify that all previous approaches are captured in a way very similar to that for complete diagnoses. This is essentially a consequence of the theorem that relates the number of consistent models for (SD, OBS, J_S) to the marginalized probability of J_S (Theorem 2). Instead of presenting the proofs again (and making repetitive arguments), we choose to allude to another set of characterizations mostly associated with consistency-based diagnosis. These are the notions of *partial* (a different characterization in consistency-based diagnosis), *kernel* and *prime* diagnoses. These notions have the same kind of drawbacks associated with the general consistency-based framework [de Kleer *et al.*, 1992] and our investigation into these notions is just in the spirit of understanding their relationship to model counting.

Definition An $AB-$literal is $AB(c)$ or $\neg AB(c)$ for some component c in $COMPS$. An $AB-$clause is a disjunction of $AB-$literals containing no complementary pair of $AB-$literals.

Definition A *conflict* of $(SD, COMPS, OBS)$ is an $AB-$clause entailed by $SD \cup OBS$. A *minimal conflict* of $(SD, COMPS, OBS)$ is a conflict no proper sub-clause of which is a conflict of $(SD, COMPS, OBS)$.

Definition *(Consistency-Based Characterization)* The *partial diagnoses* of $(SD, COMPS, OBS)$ are the implicants of the minimal conflicts of $(SD, COMPS, OBS)$.

Theorem 7 A *partial diagnosis* in the consistency-based framework identifying an implicant T of the minimal conflicts of $SD \cup OBS$, is also a *partial diagnosis* in the model-counting framework maximizing $M(J_S) = \#(SD, OBS, J_S)$ for $S =$ variables of the implicant T, but

with free variables limited to *abnormality* (AB) variables.

Proof The implicant T fixes an assignment for the components in S but leaves $COMPS \setminus S$ unassigned. Let the set of minimal conflicts of $SD \cup OBS$ be π. Let $\#_{AB}(E)$ denote the number of consistent models of E restricted to free variables being from the uninstantiated $AB-$variables. Since T is an implicant of π, all models of T (restricted to $AB-$variables) also satisfy π and are hence consistent with $SD \cup OBS$. This makes $\#_{AB}(SD, OBS, T) = \#_{AB}(T)$. In general, since $\#_{AB}(SD, OBS, T)$ is upper bounded by $\#_{AB}(T)$, the truth of the theorem follows.

Definition *(Consistency-based Characterization)* A *kernel diagnosis* identifies the prime implicants of the minimal conflicts of $SD \cup OBS$.

Without a detailed discussion (due to lack of space), we claim that this notion is related to yet another task in diagnosis — that of "representing" complete diagnoses. This task is orthogonal to "characterizing" them [Kumar, 2002]. There are other notions of diagnosis called *prime diagnoses*, *irredundant diagnoses* etc. [de Kleer *et al.*, 1992] arising mostly out of the task of "representation" and all of which are captured in one or the other way by the model counting framework (which we omit in this paper).

6 Related Work on Characterizing Diagnoses and Model Counting

Related work in trying to unify model-based and probabilistic approaches can be found in [Poole, 1994], [Kohlas *et al.*, 1998], [Lucas, 1998] and [Lucas, 2001]. [Poole, 1994] links abductive reasoning and Bayesian networks and general diagnostic reasoning systems with assumption-based reasoning. [Kohlas *et al.*, 1998] shows how to take results obtained by consistency based reasoning systems into account when computing a posterior probability distribution conditioned on the observations (the independence assumptions are lifted in [Lucas, 2001]). [Lucas, 1998] gives a semantic analysis of different diagnosis systems using basic set theory. The issue of the *modeling incompleteness assumption* is referred to in [Console *et al.*, 1989].

Diagnosis algorithms based on model counting have not yet been developed. However, the problem of model counting itself has been extensively dealt with. Although this problem is $\#P$-complete, there are a variety of techniques that have been used to make it feasible in practice (including approximate counting algorithms running in polynomial time, structure-based techniques etc.). Model counting for a SAT instance in DNF (disjunctive normal form) is simpler than it is for CNF (conjunctive normal form). For DNF, there is a fully polynomial randomized approximation scheme (FPRAS) to estimate the number of solutions [Karp *et al.*, 1989]. CDP and DDP are two model-counting algorithms for SAT instances in CNF [Bayard and Pehoushek, 2000]. A version of RELSAT has also been used to do model counting on SAT instances in CNF. If a propositional theory is in a special form called the smooth, deterministic, decomposable, negation normal form (sd-DNNF), then model counting can be made tractable and incremental [Darwiche, 2001].

7 Summary and Future Work

In this paper, we provided a unifying characterization of diagnoses based on the idea of model counting. In the process, we compared and contrasted our formalization with the previous approaches — in many cases, removing the problems associated with them. Because model counting bridges the gap between probabilities and constraints and is well-defined for many representational forms of information available about the system, we believe that the model counting characterization of diagnoses is useful and general in the sense of not imposing any restrictions on the representational form of the system description.

As for our future work, we are in the process of investigating and developing computationally tractable algorithms based on the model counting characterization of diagnoses. Advances in model counting algorithms (approximate counting, structure-based methods etc.) seem to be encouraging towards this goal. We are also working on variants of the diagnosis problem (e.g. when we are interested in a set of candidate hypotheses rather than just one).

References

[Bayard and Pehoushek, 2000] Bayard R. J. and Pehoushek J. D. Counting Models using Connected Components. *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI 2000)*.

[Console *et al.*, 1989] Console L., Theseider D., and Torasso P. A Theory of Diagnosis for Incomplete Causal Models. *Proceedings of the 10th International Joint Conference on Artificial Intelligence, Los Angeles, USA (1989) 1311-1317*.

[Console and Torasso, 1991] Console L. and Torasso P. A Spectrum of Logical Definitions of Model-Based Diagnosis. *Computational Intelligence 7(3): 133-141*.

[Darwiche, 2001] Darwiche A. On the Tractable Counting of Theory Models and its Applications to Belief Revision and Truth Maintenance. *To appear in Journal of Applied Non-Classical Logics*.

[Dechter, 1992] Dechter R. Constraint Networks. *Encyclopedia of Artificial Intelligence, second edition, Wiley and Sons, Pages: 276-285, 1992*.

[de Kleer, 1986] de Kleer J. An Assumption Based TMS. *Artificial Intelligence 28 (1986)*.

[de Kleer *et al.*, 1992] de Kleer J., Mackworth A. K., and Reiter R. Characterizing Diagnoses and Systems. *Artificial Intelligence 56 (1992) 197-222*.

[de Kleer and Williams, 1987] de Kleer J. and Williams B. C. Diagnosing Multiple Faults. *Artificial Intelligence, 32:100-117, 1987*.

[de Kleer and Williams, 1989] de Kleer J. and Williams B. C. Diagnosis with Behavioral Modes. *In Proceedings of IJCAI'89. Pages: 104-109*.

[Forbus and de Kleer, 1992] Forbus K. D. and de Kleer J. Building Problem Solvers. *MIT Press, Cambridge, MA, 1992*.

[Hamscher *et al.*, 1992] Hamscher W., Console L., and de Kleer J. Readings in Model-Based Diagnosis. *Morgan Kaufmann, 1992.*

[Karp *et al.*, 1989] Karp R., Luby M., and Madras N. Monte-Carlo Approximation Algorithms for Enumeration Problems. *Journal of Algorithms 10 429-448. 1989.*

[Kohlas *et al.*, 1998] Kohlas J., Anrig B., Haenni R., and Monney P. A. Model-Based Diagnosis and Probabilistic Assumption-Based Reasoning. *Artificial Intelligence, 104 (1998) 71-106.*

[Kumar, 2001] Kumar T. K. S. QCBFS: Leveraging Qualitative Knowledge in Simulation-Based Diagnosis. *Proceedings of the Fifteenth International Workshop on Qualitative Reasoning (QR'01).*

[Kumar, 2002] Kumar T. K. S. An Information-Theoretic Characterization of Abstraction in Diagnosis and Hypothesis Selection. *Proceedings of the Fifth International Symposium on Abstraction, Reformulation and Approximation (SARA 2002).*

[Kurien and Nayak, 2000] Kurien J. and Nayak P. P. Back to the Future for Consistency-Based Trajectory Tracking. *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI'00).*

[Lucas, 1998] Lucas P. J. F. Analysis of Notions of Diagnosis. *Artificial Intelligence, 105(1-2) (1998) 293-341.*

[Lucas, 2001] Lucas P. J. F. Bayesian Model-Based Diagnosis. *International Journal of Approximate Reasoning, 27 (2001) 99-119.*

[McIlraith, 1998] McIlraith S. Explanatory Diagnosis: Conjecturing Actions to Explain Observations. *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98).*

[Mosterman and Biswas, 1999] Mosterman P. J. and Biswas G. Diagnosis of Continuous Valued Systems in Transient Operating Regions. *IEEE Transactions on Systems, Man, and Cybernetics, 1999. Vol. 29, no. 6, pp. 554-565, 1999.*

[Nayak and Williams, 1997] Nayak P. P. and Williams B. C. Fast Context Switching in Real-time Propositional Reasoning. *In Proceedings of AAAI-97.*

[Poole, 1990] Poole D. A Methodology for Using a Default and Abductive Reasoning System. *International Journal of Intelligent Systems 5(5) (1990) 521-548.*

[Poole, 1993] Poole D. Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence, 64(1) (1993) 81-129.*

[Poole, 1994] Poole D. Representing Diagnosis Knowledge. *Annals of Mathematics and Artificial Intelligence 11 (1994) 33-50.*

[Raiman, 1989] Raiman O. Diagnosis as a Trial: The Alibi Principle. *IBM Scientific Center (1989).*

[Reiter, 1987] Reiter R. A Theory of Diagnosis from First Principles. *Artificial Intelligence 32 (1987) 57-95.*

[Shanahan, 1993] Shanahan M. Explanation in the Situation Calculus. *In Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI-93), 160-165.*

[Struss, 1988] Struss P. Extensions to ATMS-based Diagnosis. *In J. S. Gero (ed.), Artificial Intelligence in Engineering: Diagnosis and Learning, Southampton, 1988.*

[Struss and Dressler, 1989] Struss P. and Dressler O. "Physical Negation" - Integrating Fault Models into the General Diagnosis Engine. *In Proceedings of IJCAI-89. Pages: 1318-1323.*

[Williams and Nayak, 1996] Williams B. C. and Nayak P. P. A Model-Based Approach to Reactive Self-Configuring Systems. *In Proceedings of AAAI-96. Pages: 971-978.*