A Theory of Meta-Diagnosis: Reasoning about Diagnostic Systems

Nuno Belard [1,2,3] nuno.belard@airbus.com
Yannick Pencolé [2,3] ypencole@laas.fr
Michel Combacau [2,3] combacau@laas.fr

[1] Airbus France; 316 route de Bayonne; 31060 Toulouse, France
[2] LAAS-CNRS; 7 Avenue du Colonel Roche; F-31077 Toulouse, France
[3] Université de Toulouse; UPS, INSA, INP, ISAE; LAAS; 7 Avenue du Colonel Roche, F-31077 Toulouse, France

Abstract

In Model-Based Diagnosis, a diagnostic algorithm is typically used to compute diagnoses using a model of a real-world system and some observations. Contrary to classical hypothesis, in real-world applications it is sometimes the case that either the model, the observations or the diagnostic algorithm are abnormal with respect to some required properties; with possibly huge economical consequences. Determining which abnormalities exist constitutes a meta-diagnostic problem. We contribute, first, with a general theory of meta-diagnosis with clear semantics to handle this problem. Second, we propose a series of typically required properties and relate them between themselves. Finally, using our meta-diagnostic framework and the studied properties and relations, we model and solve some common meta-diagnostic problems.

1 Introduction

Diagnostic reasoning consists in determining the normal and abnormal components of a real-world system under study. In model-based diagnosis from first principles [Reiter, 1987][de Kleer and Williams, 1987] a diagnostic algorithm computes diagnoses using a model of the real-world system and some observations gathered from it. Let us, in this paper, call this model *believed system* [1]; and the tuple (believed system, observations, diagnostic algorithm) *diagnostic system*.

Now, contrary to the classical assumptions, it is ubiquitous in real-life applications for diagnostic systems to be abnormal with respect to some required properties, such as, for instance, the correctness of believed systems. At Airbus, for example, warranties that believed systems are an ontological true representation of the real-world system are needed; since not having such property would possibly result in incorrect component replacements and delays, with important economical consequences. However, having an ontologically true believed system is not always the case, and engineers struggle to find the means to detect and repair falsehoods in believed systems.

[1] The word "model" is reserved for a model-theoretic context

Let us name the problem of determining abnormalities in diagnostic systems a *meta-diagnostic* problem. Our first contribution, in Section 3, is providing Artificial Intelligence with a theory of meta-diagnosis to solve such problems. This theory has many advantages: first, it enjoys the clear semantics provided by logic; second, it is a general unified theory to reason about any model-based diagnostic system; and third it can be mapped into a theory of diagnosis and, as so, the arsenal of tools already developed through the years in the diagnostic world can be used in the meta-diagnostic one. Moreover, our theory of meta-diagnosis makes it possible to use a series of test cases (sets of observations about diagnostic systems) to refine meta-diagnoses. This is especially useful at Airbus where the data coming from many test flights can be used, for instance, to automatically achieve a perfect isolation of abnormalities in believed systems.

By making use of our theory of meta-diagnosis we contribute, in Section 5, by modelling and solving two common meta-diagnostic problems; thus providing an illustration of the developed work's potential. To do so, we also contribute, in Section 4, with a series of typically required properties of diagnostic systems and diagnoses.

2 Preliminaries

Throughout this paper, we assume that the reader is familiar with the notions of model theory (structure, model and extensions) [Hodges, 1993]. We also presume the reader familiar with the model-based diagnosis framework described by [Reiter, 1987] and [de Kleer and Williams, 1987].

2.1 Real system vs Believed System

Model-based diagnosis (MBD) is a reasoning problem that aims at retrieving system abnormalities given a system description (the so-called SD) and a set of observations OBS. For such a couple $(\mathsf{SD},\mathsf{OBS})$, the crucial assumption of MBD is that $(\mathsf{SD},\mathsf{OBS})$ *matches* with the underlying reality (i.e. the real system and the real observations). More formally speaking, reality is only accessible through a structure, let us say Ψ, of raw information everyone would have access to (in terms of behaviour and observations) if engineering and computational resources were unlimited; and the principle of model-based diagnosis is that there always exists a structure s, model of $\mathsf{SD}\cup\mathsf{OBS}$ (i.e. $\mathsf{s}\in\mathsf{Mod}(\mathsf{SD}\cup\mathsf{OBS})$), that can be extended to a structure t, i.e. $\mathsf{s}\subseteq\mathsf{t}$, which is isomorphic to

Ψ, denoted $t \rightleftharpoons \Psi$. In Tarki's terms [Tarski, 1936], MBD relies on the fact that SD$\cup$OBS is an *ontologically true* theory "which says that the state of affairs is so and so, and the state of affairs is indeed so and so". In this paper, we call *real system* a set R of interacting Replaceable Units (RU), where maintenance actions resulting from diagnosis take place. Its representation, denoted SD, will be called the *believed system*. Finally, as written above, MBD aims at retrieving abnormalities in the system, abnormalities that are always dependent on the user viewpoint on the real system:

Definition 1 (Normality and abnormality of replaceable units). *A unit $c \in R$ is said to be* abnormal *if it has passed its elastic limit and is deformed irreversibly from the standpoint of the system's user, i.e. if it cannot return to its original state in the presence of the original stimuli; and normal otherwise.*

2.2 Model-based Diagnosis

Following the notions and notations described in the previous section, we briefly recall the classical model-based diagnosis framework that will be used throughout this paper [de Kleer and Williams, 1987] [Reiter, 1987].

Definition 2 (Believed system). *A believed system $\mathcal{S}$ is a pair (SD,COMPS) where:*

1. *SD, the believed system description, is a set of first-order sentences.*
2. *COMPS, the believed system components, is a finite set of constants.*

We assume that there is a bijection between R and COMPS as R is defined by the user and the actions that can be performed on R if one of the unit is abnormal. The predicate Ab(c) (resp. $\neg$Ab(c)) represents the abnormality (resp. normality) of the component $c \in$ COMPS.

Observations are one of the few connections between real and believed systems. Intuitively, observations are captured *from the real system* by a set O of *sensors* measuring the value v(p) of a real parameter p; and used, along with the believed system and the way p and v(p) are represented, in the diagnostic reasoning.

Definition 3 (Observations). *The set of observations, OBS, is a set of first-order sentences.*

The following example illustrating a real system, its believed system counterpart and some observations will serve as a basis for some discussions throughout the paper:

Example 1. *Consider the classic circuit of Figure 1 introduced by Davis in [Davis, 1984].*

Suppose that this circuit is represented by a believed system with COMPS = $\{M_1, M_2, M_3, A_1, A_2\}$ and the SD below [2]*:*

M_1*desc:* $\neg Ab(M_1) \Rightarrow (v(x) = (v(a) + 1) * v(c))$
M_2*desc:* $\neg Ab(M_2) \Rightarrow (v(y) = v(b) * v(d))$
M_3*desc:* $\neg Ab(M_3) \Rightarrow (v(z) = v(c) * v(e))$
A_1*desc:* $\neg Ab(A_1) \Rightarrow (v(f) = v(x) + v(y))$
A_2*desc:* $\neg Ab(A_2) \Rightarrow (v(g) = v(y) + v(z))$

extended with the appropriate axioms for arithmetic and so on. Assume also that parameters a,b,c,d,e,f and g are observed and their values are 1,2,3,4,5,11 and 22 respectively.

[2]Note that the first sentence in SD is false since it does not truly represent the multiplier in the real system.

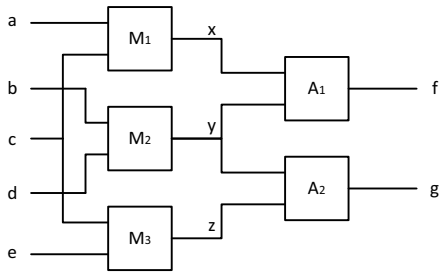

Figure 1: A circuit with 3 multipliers (M_1, M_2 *and* M_3*) and 2 adders (*A_1 *and* A_2*).*

Having a believed system (SD,COMPS) and observations OBS one can introduce the notion of *diagnostic problem*.

Definition 4 (Diagnostic problem). *A diagnostic problem DP is a tuple (SD,COMPS,OBS).*

And, finally, comes the model-theoretical definition of diagnosis relying on the concept of believed system health state, $\sigma(\Delta, \text{COMPS}\setminus\Delta) = [\bigwedge_{c \in \Delta} \text{Ab}(c)] \wedge [\bigwedge_{c \in (\text{COMPS}\setminus\Delta)} \neg \text{Ab}(c)]$.

Definition 5 (Diagnosis[3]**).** *Let $\Delta \subseteq$ COMPS. A diagnosis, D, for the diagnostic problem (SD,COMPS,OBS) is the set of all diagnosis hypotheses $\sigma(\Delta, COMPS\setminus\Delta)$ such that:*

$$SD \cup OBS \cup \sigma(\Delta, COMPS\setminus\Delta)$$

is satisfiable.

Definition 5 provides clear semantics to the notion of diagnosis. However, since it is given in terms of all models, it has little computational interest. Proof-theory comes to our rescue with a syntactic approach to logic, and hence an attractive computational environment. This is why we consider that there is a theorem-prover underlying every *diagnostic algorithm* $\mathcal{A}$. For example, in an explicit manner, Reiter in [Reiter, 1987] uses a theorem-prover in the algorithm he proposes. Model-theoretic and proof-theoretic diagnoses, are distinguished as follows:

Definition 6 (model-theoretic and proof-theoretic diagnoses).

- *A model-theoretic diagnosis, $D_{M\text{-}T}$, is the set of diagnostic hypotheses respecting Definition 5.*
- *A proof-theoretic diagnosis, $D_{P\text{-}T}$, is the set of diagnostic hypotheses computed by a diagnostic algorithm $\mathcal{A}$ ($\vdash_{\mathcal{A}}$).*

3 Characterising meta-diagnoses

Section 2 was developed around diagnostic systems.

Definition 7 (Diagnostic system). *A diagnostic system is a tuple (SD,COMPS,OBS,*$\mathcal{A}$*).*

As discussed in the introductory section, diagnostic systems can themselves be abnormal for a large number of reasons. For example: modelling errors can result in false believed systems; observations can be different from the real parameter values due to perception errors; and diagnostic algorithms can produce diagnostic hypotheses not respecting

[3]The definition of diagnosis presented corresponds to the Definition 3 by de Kleer, Mackworth and Reiter in [de Kleer *et al.*, 1992]

the model-theoretic definition. Since such abnormalities are ubiquitous in real-world applications we need a theory of meta-diagnosis for reasoning about diagnostic systems. To obtain a general theory we rely on first-order logics [4].

3.1 Meta-systems

At a meta-diagnostic level, *meta-components* can be seen as the elements of the diagnostic system whose normal or abnormal behaviour one wants to judge. This behaviour of meta-components, as well as their interactions, is described in a meta-system thanks to the unary predicate M-Ab$(\cdot)$ which carries the semantic of meta-component's abnormality. As so, intuitively, meta-systems are the static knowledge used at meta-diagnostic level to reason about diagnostic systems.

Definition 8 (Meta-system). *A meta-system is a pair (M-SD,M-COMPS) where:*

1. *M-SD, the meta-system description, is a set of first-order sentences.*
2. *M-COMPS, the meta-system components, is a finite set of constants.*

The choice of meta-components depends on our goals and underlying hypotheses. As so:

1. It is not mandatory for every element of a diagnostic system to be considered as a meta-component.
2. Meta-components can be defined at different abstraction levels.

For instance, if for a given problem one assumes that diagnostic algorithms and observations are never abnormal; to determine if each sentence in the believed system description describes a real system behaviour in a correct manner; one can associate a meta-component to every sentence in the believed system description and associate no meta-components to the rest of the diagnostic system. This is exactly what we illustrate below based on the diagnostic system of Example 1:

Example 1 (continued). *One possibility for M-SD with M-COMPS=$\{M_1desc, M_2desc, M_3desc, A_1desc, A_2desc, Alg\}$ would then be [5]:*

$$\neg M\text{-}Ab(M_1desc) \Rightarrow [\neg Ab(M_1) \Rightarrow (v(x) = (v(a)+1)*v(c))]$$
$$\neg M\text{-}Ab(M_2desc) \Rightarrow [\neg Ab(M_2) \Rightarrow (v(y) = v(b) * v(d))]$$
$$\neg M\text{-}Ab(M_3desc) \Rightarrow [\neg Ab(M_3) \Rightarrow (v(z) = v(c) * v(e))]$$
$$\neg M\text{-}Ab(A_1desc) \Rightarrow [\neg Ab(A_1) \Rightarrow (v(f) = v(x) + v(y))]$$
$$\neg M\text{-}Ab(A_2desc) \Rightarrow [\neg Ab(A_2) \Rightarrow (v(g) = v(y) + v(z))]$$

The idea behind such M-SD is that if SD-sentence is not abnormal, then what the sentence describes happens in reality.

3.2 Meta-observations

Meta-observations are used along with meta-systems to determine the normal and abnormal meta-components.

Definition 9 (Meta-observations). *The set of meta-observations, M-OBS, is a finite set of first-order sentences.*

Examples of meta-observations obtained from available test cases could be real system health state σ_{real}, the diagnoses computed by a diagnostic algorithm, observations and so on. In fact, meta-observations can be every possible observation at a diagnosis level along with every possible observation about the diagnostic system itself.

Example 1 (continued). *Imagine the real system health state is known for the OBS in this example. Then, M-OBS is:*

$$\sigma_{real}: \quad \neg Ab(M_1)\wedge\neg Ab(M_2)\wedge\neg Ab(M_3)\wedge \wedge\neg Ab(A_1)\wedge Ab(A_2)$$
$$OBS: \quad v(a)=1\wedge v(b)=2\wedge v(c)=3\wedge v(d)=4\wedge \wedge v(e)=5\wedge v(f)=11\wedge v(g)=22$$

3.3 Meta-diagnoses

Meta-systems and meta-observations can be grouped in a meta-diagnostic problem.

Definition 10 (Meta-diagnostic problem). *A meta-diagnostic problem M-DP is a tuple (M-SD,M-COMPS,M-OBS).*

A meta-diagnostic problem is now ready to be solved. Doing so consists in determining the normality or abnormality of meta-components M-COMPS, based on meta-observations.

Definition 11 (Meta-health state). *Let $\Phi \subseteq$ M-COMPS be a set of meta-components. The meta-health state $\pi(\Phi,$M-COMPS$\setminus\Phi)$ is the conjunction:*

$$[\bigwedge_{mc\in\Phi} M\text{-}Ab(mc)]\wedge[\bigwedge_{mc\in(M\text{-}COMPS\setminus\Phi)} \neg M\text{-}Ab(mc)]$$

From the notion of meta-health state one can move to solving the meta-diagnostic problem.

Definition 12 (Meta-diagnosis). *Let $\Phi \subseteq$ M-COMPS. A meta-diagnosis, M-D, for the meta-diagnostic problem (M-SD,M-COMPS,M-OBS) is the set of all meta-diagnostic hypotheses $\pi(\Phi,$M-COMPS$\setminus\Phi)$ such that:*

$$M\text{-}SD \cup M\text{-}OBS \cup \pi(\Phi, M\text{-}COMPS\setminus\Phi)$$

is satisfiable.

All in all, the reader may notice that Definitions 5 and 12 are perfectly equivalent. Hence, a meta-diagnostic problem can be seen as a diagnostic problem where the artefact being diagnosed is a diagnostic system. There are numerous advantages to this analogy. We highlight the following:

1. Every diagnostic algorithm can become a meta-diagnostic algorithm if it is sound and complete with respect to the underlying semantics [Hodges, 1993].
2. Every approach to handle the complexity problems of diagnosis can be used in meta-diagnosis.

4 Defining properties of diagnostic systems and diagnostic results

In the last section we proposed a characterisation of meta-diagnoses general enough to handle many different meta-diagnostic problems. Now, we introduce some usually required properties of diagnostic systems and diagnostic results; whose absence is considered abnormal. In Section 5 we rely on such properties to model some typical meta-diagnostic problems. We encourage the reader to refer, for example, to [Belard *et al.*, 2010] for some more properties.

[4] The results obtained are also valid for a series of other logics.

[5] Section 5 clarifies the rational for meta-system description.

4.1 Diagnostic result properties

Model-theoretic and proof-theoretic diagnoses' quality can be evaluated thanks to two properties: *validity* and *certainty*.

Definition 13 (Validity of a diagnosis). *Let σ_{real} be the believed system health state such that, for every $c \in COMPS$, if c is the image of $r \in R$: 1) if r is abnormal, $\neg Ab(c) \wedge \sigma_{real} \models \bot$; and 2) if r is normal, $Ab(c) \wedge \sigma_{real} \models \bot$. A diagnostic result, D, is said to be valid if $\sigma_{real} \in D$; and invalid otherwise.*

Definition 14 (Certainty of a diagnosis). *A diagnostic result, D, is said to be certain if there is a single diagnosis hypothesis, i.e. $\#D = 1$; and uncertain otherwise.*

Having valid and certain diagnoses is extremely interesting for most real life diagnostic applications. In the aeronautic industry, for example, invalid diagnoses may lead the aircraft maintenance team to replace the wrong components; and uncertain diagnoses naturally increase the time of repair.

4.2 Observations properties

Let us focus on a single property of observations: truth. Informally, true observations assure a correct perception of the real parameter values; and if the value of a given parameter p is observed to be x then we know that in reality the parameter p has the value x. More formally:

Definition 15 (Truth of the observations). *Let Ω be the set of all structures and $\Psi \in \Omega$ the raw information about the reality* [6]*. The observations OBS are an ontological truth iff $\exists_{s \in Mod(OBS)} \exists_{t \in \Omega} (s \subseteq t) \wedge (t \rightleftharpoons \Psi)$ [Tarski, 1936]. If OBS is an ontological truth then so are every sentences in OBS.*

Note that there is a difference between ontological and logical truths; for the latter are the axioms in a theory, while the former establishes a correspondence between the sentences of the theory and reality [Tarski, 1936]. Without the ontological truth of observations the model-theoretic diagnoses are not guaranteed to be valid as we will prove later on.

4.3 Believed system properties

Hereafter we define two properties one usually wants believed systems to have: truth and diagnosability.

Let us start with the property of truth of believed systems. Intuitively, if a believed system is ontologically true, then if a sentence states X, X happens in reality. More formally:

Definition 16 (Truth of the believed system). *Let Ω be the set of all structures and $\Psi \in \Omega$ the raw information about the reality. A believed system is an ontological truth iff, for all true OBS, $\exists_{s \in Mod(SD \cup OBS)} \exists_{t \in \Omega} (s \subseteq t) \wedge (t \rightleftharpoons \Psi)$. If a believed system is an ontological truth then so are every SD-sentences.*

In the same way it is important to have ontological true observations, i.e. a correct perception of the real parameter values, the same goes with the ontological truth of believed systems; since without this property the model-theoretic diagnoses are not guaranteed to be valid as we will prove later.

As for the diagnosability property of believed systems, let us borrow its definition from [Console *et al.*, 2000] and rephrase it to better suit our framework as follows:

[6] For a more detailed discussion cf. Subsection 2.1.

Definition 17 (Diagnosability of the believed system). *A believed system with a set of sensors O is said to be diagnosable iff for any ontologically true observations there is always one and only one model-theoretic diagnosis hypothesis.*

The interest of the diagnosability is naturally justified by the need for certain diagnoses.

4.4 Diagnostic algorithm properties

Being a theorem prover, a diagnostic algorithm can enjoy from soundness and completeness properties:

Definition 18 (Soundness and completeness of the diagnostic algorithm). *Let T and φ be, respectively, a logical theory and a sentence in a language $\mathcal{L}$ with a given semantics. A diagnostic algorithm $\mathcal{A}$ is sound iff:*

$$\text{If } (T \vdash_{\mathcal{A}} \varphi), \text{ then } (T \models_{\mathcal{L}} \varphi)$$

and complete iff:

$$\text{If } (T \models_{\mathcal{L}} \varphi), \text{ then } (T \vdash_{\mathcal{A}} \varphi)$$

One can see that semantic entailment and syntactic proof are equivalent for sound and complete diagnostic algorithms. This is why such properties are extremely interesting and, for instance, Reiter in [Reiter, 1987] explicitly requires a sound and complete theorem prover in his diagnostic algorithm. Moreover, one could question the interest of meta-diagnosing instead of directly studying the soundness and/or completeness of an algorithm. The answer comes from the black-box vision of algorithms in many real-world applications; for they are either too hard to model in logic or it is just impossible to access the details of the algorithm. An example is the centralized maintenance system diagnostic algorithm at Airbus.

5 Modelling and solving meta-diagnostic problems

With the characterisation of meta-diagnosis of Section 3 and the properties of Section 4 we can model and solve some typical meta-diagnostic problems; thus providing the reader with a flavour of what can be done with such framework.

We propose the following steps for the modelling process:

1. Define meta-components based on the problem hypotheses and the detail level wanted for the solution.
2. Define properties whose absence is abnormal.
3. Define meta-components' normal/abnormal behaviour.
4. Define meta-observations.

As for solving the meta-diagnostic problems, we have used the General Diagnostic Engine (GDE) [de Kleer and Williams, 1987][Forbus and de Kleer, 1993]. This was possible because, as discussed in Section 3, every sound and complete diagnostic algorithm can become a meta-diagnostic algorithm; which is the case of GDE. As expected, it was enough to add the definitions of the meta-system and meta-observations to GDE without changing its core modules.

5.1 The problem of false believed systems

Imagine we are given the following problem, typically when testing if a newly modelled believed system is a good representation of the real system: "In a given diagnostic system (SD,COMPS,OBS,$\mathcal{A}$) the observations are assumed to be an

ontological truth and the diagnostic algorithm is assumed to be sound and complete. The believed system, however, can be false. Find the ontologically false sentences in SD.".

The following theorem is needed to handle this problem:

Theorem 1. *If (SD,COMPS) is an ontologically true believed system, then for every diagnostic problem (SD,COMPS,OBS) with ontologically true observations, every model-theoretic diagnosis $D_{M\text{-}T}$ is valid.*

Proof. Suppose σ_{real} is not a believed system health state determined using the model-theoretic Definition 5. If so, $\text{SD} \cup \text{OBS} \cup \sigma_{\text{real}}$ is unsatisfiable, thus having no model. Now, σ_{real}, coming from the real system, must have a model s *such that* $\exists_{t \in \Omega}\ (s \subseteq t) \wedge (t \rightleftharpoons \Psi)$. *Combining all the arguments we get that* $\neg(\exists_{s \in Mod(SD \cup OBS)}\ \exists_{t \in \Omega}\ (s \subseteq t) \wedge (t \rightleftharpoons \Psi))$. *As so, either the believed system or the observations are not true. Q.E.D.* □

Imagine an instance of this problem where the SD, COMPS and OBS are the ones from Example 1.

Now, this is a typical meta-diagnostic problem. Let us follow the proposed modelling process:

1. The meta-components are the sentences in SD.
2. The property whose absence is considered abnormal is the ontological truth of each sentence.
3. Suppose an SD-sentence stating "A" represented by a meta-component mc_1. Now, either A is ontologically true or mc_1 is abnormal, i.e. $\text{M-Ab}(\text{mc}_1) \vee \text{A}$. So, the normal behaviour of this sentence would be described, at a meta-system level, as: $\neg\text{M-Ab}(\text{mc}_1) \Rightarrow \text{A}$.
4. From Theorem 1 and the problem hypotheses we get that if every meta-component is normal then the model-theoretic diagnosis is valid. To determine the validity of the model-theoretic diagnosis we need to meta-observe the real health state and the observations [7].

Following such modelling process in our problem instance resulted in the M-SD and M-COMPS also given in Example 1. If M-OBS are also those of that example we get the following meta-diagnosis (from GDE output):

There are 3 minimal [kernel] diagnoses:
$\{\text{M}_1\text{desc}\}$; $\{\text{M}_2\text{desc}\}$; $\{\text{A}_1\text{desc}\}$

Although not certain, the meta-diagnosis correctly includes the ontological falsehood of the sentence M_1desc.

5.2 The problem of diagnosable believed systems and sound and complete diagnostic algorithms

Suppose we corrected the M_1desc sentence using the meta-diagnosis computed in the last subsection; and we can now hypothesise that the believed system is true. Imagine some further requirements, typically when testing a new diagnostic algorithm and considering it as a black box, as well as exploring further properties of the believed system: "Admit the ontological truth of the believed system and the possible lack of soundness and/or completeness by the diagnostic algorithm. Find if the believed system with sensors O is diagnosable and if the diagnostic algorithm is sound and complete".

Once again, let us first introduce three theorems which are needed when handling this problem:

Lemma 2. *If $\mathcal{A}$ is a sound and complete diagnostic algorithm, then proof-theoretic diagnoses and model-theoretic diagnoses are equivalent.*

Proof. The model-theoretic definition of diagnosis (cf. Definition 5) states that $\sigma(\Delta,\text{COMPS}\backslash\Delta)$ *is a diagnostic hypothesis iff* $\text{SD} \cup \text{OBS} \cup \sigma(\Delta,\text{COMPS}\backslash\Delta)$ *is satisfiable.*

One can make the bridge between model-theory and proof-theory shine by stating that an anti-diagnostic hypothesis is $\sigma(\Delta,\text{COMPS}\backslash\Delta)$ *such that* $\text{SD} \cup \text{OBS} \models \neg\sigma(\Delta,\text{COMPS}\backslash\Delta)$*; and stating that every* $\sigma(\Delta,\text{COMPS}\backslash\Delta)$ *that is not an anti-diagnostic hypothesis is a diagnostic hypothesis.*

Since $\mathcal{A}$ is sound and complete, model-theoretic and proof-theoretic anti-diagnostic hypotheses are the same; and since anti-diagnosis and diagnosis are complementary, model-theoretic and proof-theoretic diagnoses are equivalent for a sound and complete diagnostic algorithm. Q.E.D. □

Theorem 3. *If $\mathcal{A}$ is a sound and complete diagnostic algorithm and (SD,COMPS) is an ontologically true believed system, then for every diagnostic problem (SD,COMPS,OBS) with ontologically true observations, every proof-theoretic diagnosis, $D_{P\text{-}T}$, is valid.*

Proof. Trivial from Lemma 2 and Theorem 1. Q.E.D. □

Theorem 4. *If $\mathcal{A}$ is a sound and complete diagnostic algorithm and (SD,COMPS) is an ontological true and diagnosable believed system with sensors O, then for every diagnostic problem (SD,COMPS,OBS) with ontologically true OBS, every proof-theoretic diagnosis $D_{P\text{-}T}$ is valid and certain.*

Proof. Trivial from Definition 17 and Theorem 3. □

As an instance of this problem suppose, once again, the SD, COMPS and OBS from Example 1 but with the SD-sentence M_1desc corrected.

Let us, once more, use the proposed modelling process:

1. The meta-components in this problem are the diagnostic algorithm Alg and the system description SD_{comp}.
2. The properties whose absence is considered abnormal are the diagnosability of the believed system with sensors O and the soundness and/or completeness of the diagnostic algorithm.
3. From Theorem 3 and the problem hypotheses we get the normal behaviour of the diagnostic algorithm. If Alg is the meta-component associated to $\mathcal{A}$ we get: $\neg\text{M-Ab}(\text{Alg}) \Rightarrow (\sigma_{\text{real}} \Rightarrow \text{Dis}(\text{D}_{\text{P-T}}))$; where $\text{Dis}(\cdot)$ is a function that returns the disjunction of every element in a set. Moreover, using Theorem 4 and the problem hypotheses we get the normal behaviour for the meta-component SD_{comp}: $\neg\text{M-Ab}(\text{Alg}) \wedge \neg\text{M-Ab}(\text{SD}_{\text{comp}}) \Rightarrow (\#\text{D}_{\text{P-T}} = 1)$
4. The meta-observations are the proof-theoretic diagnoses $\text{D}_{\text{P-T}}$ and the real health state σ_{real}.

[7] In many real-world applications it is possible to observe the real health state or at least a part of it. In the aeronautic domain, for instance, we have access to the real health state since each replaced component is tested after its removal.

By following such modelling process we could obtain, for the instance of the problem from Example 1 but with the SD-sentence M_1desc corrected, M-COMPS = $\{Alg, SD_{comp}\}$ and the following M-SD (extended with the appropriate axioms for arithmetic and so on):

$$\neg\text{M-Ab}(Alg) \Rightarrow (\sigma_{real} \Rightarrow Dis(D_{P\text{-}T}))$$
$$\neg\text{M-Ab}(Alg) \wedge \neg\text{M-Ab}(SD_{comp}) \Rightarrow (\#D_{P\text{-}T} = 1)$$

Now, suppose the meta-observations are $\sigma_{real} = \neg Ab(M_1) \wedge \neg Ab(M_2) \wedge \neg Ab(M_3) \wedge \neg Ab(A_1) \wedge Ab(A_2)$ and the proof-theoretic diagnoses $D_{P\text{-}T}$ are all the diagnosis hypotheses covered by the following four kernel diagnoses: $\{A_2\}$, $\{A_1 \wedge M_2\}$, $\{M_3\}$ and $\{M_1 \wedge M_2\}$.

We get the following meta-diagnosis (from GDE output):

There are 2 minimal [kernel] diagnoses: $\{Alg\}$; $\{SD_{comp}\}$

Now, with our meta-diagnosis we detected that there is an abnormality somewhere in the meta-components, but we had very poor isolation performance. To our rescue comes the fact that meta-diagnostic reasoning is monotonic if we assume M-OBS is true, since M-SD is built upon theorems. As so, we can use another test case, i.e. another meta-observations, to refine our meta-diagnoses. Let us suppose another test case was received where the meta-observations are $\sigma_{real} = \neg Ab(M_1) \wedge \neg Ab(M_2) \wedge \neg Ab(M_3) \wedge \neg Ab(A_1) \wedge Ab(A_2)$ and the proof-theoretic diagnoses are all the diagnosis hypothesis covered by the following three kernel diagnoses: $\{A_1 \wedge M_2\}$, $\{M_3\}$ and $\{M_1 \wedge M_2\}$.

We get the following meta-diagnosis (from GDE output):

There is 1 minimal [kernel] diagnosis: $\{Alg\}$

The meta-diagnosis is, thus, refined.

6 Related work

In the model-based diagnosis community, one can find some works that recognise the existence of abnormalities in diagnostic systems, even if very locally in the whole spectrum of possible abnormalities. Examples of such recognition are Davis and Hamscher's, statement in [Davis and Hamscher, 1988]: "a model is never completely correct"; Console, Dupré and Torasso's statement in [Console *et al.*, 1989]: "[in some cases] a complete model is either not available or intractable"; or Struss's work on abstractions and simplifications of models in [Struss, 1992].

As for managing abnormalities, few attempts have been made. Exceptions are, for instance, the management of incomplete believed systems in [Console *et al.*, 1989] and [Yeung and Kwong, 2005] or the management of uncertain observations in [Lamperti and Zanella, 2002].

Finally, to our best knowledge, in the model-based diagnostic community there has been little work done on detecting and isolating abnormalities in diagnostic systems; despite the interest and ubiquity of the problem. In fact, the closer we can get to such results is the work of Yeung and Kwong in [Yeung and Kwong, 2005] where the authors not only focus on fault detection and isolation but also attempt to learn what to change to repair the believed system incompletenes.

All in all, our theory of meta-diagnosis provides model-based diagnosis with a uniform framework for dealing with possibly abnormal diagnostic system. Moreover, meta-diagnosis can be used to isolate abnormalities in diagnostic systems and select the best approach to manage them. For example, if a meta-diagnosis is the incompleteness of a believed system, then one can deal with such incompleteness by using Console, Dupré and Torasso's approach.

7 Conclusions and perspectives

In this paper we have proposed a general theory of meta-diagnosis and have modeled and solved two common meta-diagnostic problems; by making use of some properties of diagnostic systems and diagnostic results we proposed. By showing that the meta-diagnostic task can be transformed into a diagnostic task we have proved that diagnostic-world tools such as algorithms or techniques to manage computational complexity can be used at a meta-diagnostic level. GDE usage in Section 5 illustrates this point. The meta-diagnostic framework proposed can be used either to validate the classical hypothesis of no abnormalities in diagnostic systems; or, if some abnormalities are present, to choose the best approach in the model-based diagnosis community to cope with them.

Our work in meta-diagnosis opens the doors to numerous applications, both in the industrial and academic worlds; some of which have already been successfully implemented. The automatic detection and isolation of errors in the Centralized Maintenance System's knowledge base in aircraft or the automatic validation of diagnostic algorithms seen as black-boxes are just some of the potential usages of our results.

In the future we plan to: 1) focus on studying the application of meta-diagnosis to systems other that diagnostic ones, such as planning or prognostic models; 2) instantiating a meta-diagnoser at Airbus to improve the quality of the centralized maintenance system; and 3) focus on automatic repair of abnormal meta-components (eg. restoring the ontological truth of believed system sentences by learning their connection with some hidden variables).

References

[Belard *et al.*, 2010] Nuno Belard, Yannick Pencolé, and Michel Combacau. Defining and exploring properties in diagnostic systems. In *DX-10 21th International Workshop on Principles of Diagnosis*, 2010.

[Console *et al.*, 1989] Luca Console, Daniele Theseider Dupré, and Pietro Torasso. A theory of diagnosis for incomplete causal models. In *Proc. 11th IJCAI*, 1989.

[Console *et al.*, 2000] Luca Console, Claudia Picardi, and Marina Ribaudo. Diagnosis and diagnosability analysis using pepa. In *ECAI*, 2000.

[Davis and Hamscher, 1988] Randall Davis and Walter Hamscher. Model-based reasoning: Troubleshooting. Technical report, M.I.T., 1988.

[Davis, 1984] Randall Davis. Diagnostic reasoning based on structure and behavior. *Artif. Intell.*, 24(1-3), 1984.

[de Kleer and Williams, 1987] Johan de Kleer and Brian C. Williams. Diagnosing multiple faults. *Artif. Intell.*, 32(1), 1987.

[de Kleer *et al.*, 1992] Johan de Kleer, Alan K. Mackworth, and Raymond Reiter. Characterizing diagnoses and systems. *Artif. Intell.*, 56(2-3), 1992.

[Forbus and de Kleer, 1993] Kenneth D. Forbus and Johan de Kleer. *Building Problem Solvers*. M.I.T. University Press, 1993.

[Hodges, 1993] Wilfrid Hodges. *Model Theory*. Number 42 in Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1993.

[Lamperti and Zanella, 2002] Gianfranco Lamperti and Marina Zanella. Diagnosis of discrete-event systems from uncertain temporal observations. *Artif. Intell.*, 137, 2002.

[Reiter, 1987] Raymond Reiter. A theory of diagnosis from first principles. *Artif. Intell.*, 32(1), 1987.

[Struss, 1992] Peter Struss. *What's in SD?: Towards a theory of modeling for diagnosis*. Morgan Kaufmann Publishers Inc., 1992.

[Tarski, 1936] Alfred Tarski. The concept of truth in formalized languages. In *Logic, Semantics, Metamathematics*. Oxford University Press, Oxford, 1936.

[Yeung and Kwong, 2005] David L. Yeung and Raymond H. Kwong. Fault diagnosis in discrete-event systems: Incomplete models and learning. In *Proceedings of the 2005 American Control Conference*, volume 5, 2005.