Multi-agent Diagnosis: an analysis

Nico Roos [b] Annette ten Teije [c] André Bos [a]
Cees Witteveen [a]

[a] Delft University of Technology, ITS, P.O.Box 256, 2600 AJ Delft.
[b] Universiteit Maastricht, Infonomics / IKAT, P.O.Box 616, 6200 MD Maastricht.
[c] Utrecht University, ICS, P.O.Box 80.089, 3508 TB Utrecht.

Abstract

This paper analyzes the use of a multi-agent System for Model-Based Diagnosis. In a large dynamical system, it is often infeasible or even impossible to maintain a model of the whole system. Instead, several incomplete models of the system have to be used to detect possible faults. These models may also be physically distributed.

A Multi-Agent System of diagnostic agents may offer solutions for establishing a global diagnosis. If we use a separate agent for each incomplete model of the system, establishing a global diagnosis becomes a problem of cooperation and negotiation between the diagnostic agents. This raises the question whether '*a set of diagnostic agents, each having an incomplete model of the system, can (efficiently) determine the same global diagnosis as an ideal single diagnostic agent having the combined knowledge of the diagnostic agents?*'.

1 Introduction

A traditional diagnostic tool can be viewed as a single *diagnostic agent* having a model of the whole system to be diagnosed. There are, however, several reasons why such a single agent approach may be inappropriate. First of all, if the system is physically distributed and large, there may be not enough time to compute a diagnosis centrally and to communicate all observations. Secondly, if the structure of the system is dynamic, it may change too fast to maintain an accurate global model of the system over time. Finally, sometimes the existence of an overall model is simply undesirable. For example, if the system is distributed over different legal entities, one entity does not wish other entities to have a detailed model of its part of the system. Examples of such systems are modern telecommunication networks, dynamic configuration of robotic systems such as AGV driving in a platoon, and so on. For such systems, a *distributed* approach of multiple diagnostic agents might offer a solution.

An important question is of course whether a set of diagnostic agents can (efficiently) determine the same global diagnosis as an ideal single diagnostic agent having the combined knowledge of the diagnostic agents?

To investigate this problem we distinguish two ways in which the model (knowledge) is distributed over the agents (cf. [4]): (1) *spatially distributed*: knowledge of system behavior is distributed over the agents according the spatial distribution of the system's components, and (2) *semantically distributed*: knowledge of system behavior is distributed over the agents according to the type of knowledge, e.g. a separate model of the electrical

and of the thermodynamical behavior of the system. We will not consider approaches in which all agent use the same model [9] in order to gain fault-tolerant behavior.

The way the knowledge is distributed turns out to have significant repercussion on multi-agent diagnosis.[1]

This paper is organized as follows. Section 2 specifies the diagnostic problem and Section 3 gives the standard diagnostic definitions. Section 4 discusses multi-agent diagnosis. Section 5 concludes the paper.

2 The diagnostic setting

A system to be diagnosed is a tuple $S = (C, M, Id, Sd, Ctx, Obs)$ where C is a set of components, $M = \{M_c \mid c \in C\}$ is a specification of possible fault modes per component, Id is a set of identifiers of connection points between components, Sd is the system description, Ctx is a specification of input values of the system that are determined outside the system by the environment and Obs is a set of observed values of the system. A component in C has a normal mode $nor \in M_c$, one general fault mode $ab \in M_c$ and possibly several specific fault modes.

We assume that all components have *in*- and *outputs* and that every in- and output only has one value type; e.g.: current, voltage, temperature, and so on. This assumption is not valid in every system. We cannot say, for instance, that a resistor has an input and an output. Without the assumption, the behavioral description of a component constraints the values of certain types the connection points of a component may take. Determining the behavior of a system requires solving a distributed (non-linear) constraint satisfaction problem when using this kind of behavioral descriptions for components. Solving such a problem can be hard when the knowledge is distributed over several agents.

Note that we can transform the behavioral description of a component into an equivalent description in which the component has only in- and outputs, for instance, by using Bond Graphs [3]. In the resulting description, one physical connection point may be represented by several in- and outputs.

The system description $Sd = Str \cup Beh$ consists of a structural description Str and a behavioral description for each component $Beh = \bigcup_{c \in C} Beh_c$. The structural description Str consists of instances of the form $p = in(x, c)$ and $p = out(x, c)$ where x is an in- or an output of a component c and $p \in Id$ is a connection point identifier. Of course, a connection point $p \in Id$ has at most one output. The function $type(p)$ will be used to denote the value type of a connection point $p \in Id$.

The set Beh_c specifies a behavior for each (fault) mode in M_c of a component c, possibly with the exception of $ab \in M_c$. In this specification, the predicate $mode(c, m)$ is used to denote the mode $m \in M_c$ of a component c. For each instance of $mode(,)$, Beh_c specifies a behavioral description of the form: $mode(c, m) \rightarrow \Phi$ where $m \in M_c$.[2] The expression Φ describes the component's behaviour given its mode $m \in M_c$.

The set Ctx describes the values of connection points that are determined by the environment. It consists of instances of the form $value(p) = v$ where $p \in Id$ is a

[1] Although we distinguish spatially and semantically distributed models, combinations are also possible.

[2] Note that we may use a single description for a class of components. Instances of this description must imply the form of description give here.

connections point and v is a value.

Finally, the set Obs describes the values of connection points that are observed (measured) by the diagnostic agent. It also consists of instances of the form $value(p) = v$ where $p \in Id$ is a connection point and v is a value.

A *candidate diagnosis* is an assignment of modes that might explain the observed behavior of a system. The candidate diagnosis is a set D of instances of the predicate $mode(,)$ such that for every component $c \in C$ there is exactly one mode in $m \in M_c$ such that $mode(c, m) \in D$. A *diagnosis* is a candidate diagnosis that explains the observed behaviour of a system $S = (C, M, Sd, Ctx, Obs)$ according to our diagnostic definition, to be discussed in the next section.

Note that there can be more than one diagnosis, only one of which gives the correct explanation. The latter is called the final diagnosis.

3 Single agent diagnosis

In this section we present some well-known concepts in model-based diagnosis. It will be called *single agent diagnosis* since it assumes that a single agent, having complete knowledge of the system, $S = (C, M, Sd, Ctx, Obs)$, suffices to make a diagnosis.

The diagnostic definition Given a system $S = (C, M, Sd, Ctx, Obs)$, a diagnosis can be made. In the literature two types of diagnoses are distinguished: *consistency based* [6, 7] and *abductive* [1] diagnosis. Both can be combined into one more general diagnostic definition [2]. This definition will be used here:

Definition 1 *Let $S = (C, M, Sd, Ctx, Obs)$ be the system to be diagnosed. Let Obs_{con}, $Obs_{\text{abd}} \subseteq Obs$ be two subsets of the observations and let D be a candidate diagnosis. Then D is a diagnosis for S iff*

$$D \cup Sd \cup Ctx \mid\!\sim \bigwedge_{\varphi \in Obs_{\text{abd}}} \varphi \text{ and } D \cup Sd \cup Ctx \cup Obs_{\text{con}} \not\vdash \bot.$$

Note that we use the symbol $\mid\!\sim$ to denote the possibly limited reasoning capabilities of a diagnostic system. I.e $\{\varphi \mid \Sigma \mid\!\sim \varphi\} \subseteq \{\varphi \mid \Sigma \vdash \varphi\}$.

If $Obs_{\text{abd}} = \varnothing$, then we have a pure consistency-based diagnosis, and if $Obs_{\text{con}} = \varnothing$, we have a pure abductive diagnosis. Note that an abductive diagnosis is stronger than consistency-based diagnosis since the former also requires: $D \cup Sd \cup Ctx \cup \varnothing \not\vdash \bot$.

Besides pure consistency based and abductive diagnosis, there is another interesting special case. In the absence of fault models, usually consistency based diagnosis is used since we cannot explain *abnormal observations*; i.e. the observations that do not correspond with the predicted values in case of no component failures. We can improve consistency based diagnosis if we also allow for abductive diagnosis [8]. In the absence of fault models, we can only give an explanation for the normal observations Obs_N; i.e. the observations that correspond with the predicted values in case of no component failures. This additional information can help us to reduce the number of candidate diagnoses, especially if it is safe to assume that the effects of one fault cannot be compensated by the effects of other faults.

The number of diagnoses Potentially, there can be an exponential number of diagnoses. Even for relatively small systems, listing all these diagnoses can be infeasible. In a well designed system it is unlikely that the many components fail at the same time (unless there is a cascade of failures). So, it is safe to assume that only a minimal number of components is broken. Hence, we can order the diagnoses with respect to the number of broken components. We can look for either diagnoses with a *minimum* number or with a *subset-minimal* number of broken components. Here we choose the latter.

Definition 2 *Let D and D' be two diagnoses. D is less than D', $D \prec D'$, iff $\{c \mid mode(c, ab) \in D\} \subset \{c \mid mode(c, ab) \in D'\}$. A diagnosis D is* minimal *iff for no diagnosis D' it holds that $D' \prec D$.*

Minimal diagnoses have a property that enable them to characterize a whole set of diagnoses [5, 7]. This property turns out to be useful if we need to combine the diagnoses made by several agents:

Proposition 1 *Suppose that for each component $c \in C$ there are exactly two modes, nor and ab, and let $D \prec D'$ be two candidate diagnoses. Then D' is a pure consistency based diagnosis of a system if D is.*

This is a nice result since it enables us to characterize an exponential number of diagnoses. Especially if the number of faults is bounded by a constant or of the order $\mathcal{O}(\log(|C|))$, the number of minimal diagnoses is polynomial in $|C|$.

Partial diagnoses are another way to avoid listing an exponential number of diagnoses. In a partial diagnosis the mode of some of the components $c \in C$ is left undefined.[3]

Definition 3 *Let D' be some candidate diagnosis. Then $D \subseteq D'$ is a partial diagnosis.*

We are of course interested in the smallest set, with respect to $\subset$, of components such that the corresponding partial diagnoses characterize a set of diagnoses. This partial diagnosis is called a *kernel diagnosis* [5].

Definition 4 *D is a* kernel diagnosis *of a system iff (1) D is a partial diagnosis such that every candidate diagnosis $D' \supseteq D$ is a diagnosis of the system, and (2) for no partial diagnosis $D'' \subset D$ the first item holds.*

Definition 5 *D is an* abductive kernel diagnosis *iff D is a minimal partial diagnosis such that: $D \cup Sd \cup Ctx \mid\sim \bigwedge_{\varphi \in Obs_{\mathrm{abd}}} \varphi$.*[4]

Definition 6 *D is a* consistency based kernel diagnosis *if and only if D is a minimal partial diagnosis such that: $D \cup Sd \cup Ctx \cup Obs_{\mathrm{con}} \not\mid\sim \bot$.*[5]

We can derive the kernel diagnoses for consistency based diagnosis with abductive explanation of normal observations from the two types of kernel diagnoses defined above.

[3] Our definition of a partial diagnosis differs from the definition given in [5].

[4] Note that all mode descriptions in D have the *normal* mode if $Obs_{abd} = Obs_N$. Also note that there is only one kernel diagnosis if none of the components behaves like a switch [8].

[5] Note that without fault models all mode descriptions in D have the *abnormal* mode.

Proposition 2 *Let D^{abd} be an abductive kernel diagnosis and let D^{con} be a consistency based kernel diagnosis of a system. Then, $D = D^{abd} \cup D^{con}$ is a kernel diagnosis that characterizes consistency based diagnosis with abductive explanation of normal observations if D is a partial diagnosis.*[6]

Proposition 3 *Let D be a kernel diagnosis that characterizes consistency based diagnosis with abductive explanation of normal observations.*

Then $D^{abd} = \{mode(c, nor) \mid mode(c, nor) \in D\}$ is an abductive partial diagnosis and $D^{con} = \{mode(c, ab) \mid mode(c, ab) \in D\}$ is a consistency based kernel diagnosis.

4 Multi-agent diagnosis

Suppose that instead of one diagnostic agent, we have two or more diagnostic agents. What can we say about the ability of this group of agents to make a diagnosis. We will only consider cases in which we have two diagnostic agents since any case in which we have $n > 2$ diagnostic agents is a trivial extension. We assume that both agents, A_1 and A_2, have partial knowledge about the system. Let $C = C_1 \cup C_2$, let $Sd = Sd_1 \bigcup Sd_2$ and let $Obs = Obs_1 \bigcup Obs_2$. We also assume that each agent knows the connections with the other agent. An agent may have to ask / tell the values of these connection points from / to another agent. We divide the connection points between subsystems into inputs $In_i = \{p \in Id \mid \{p = in(x, c), p = out(y, c')\} \subseteq Str, c \in C_i, c' \notin C_i\}$ and outputs $Out_i = \{p \in Id \mid \{p = out(x, c), p = in(y, c')\} \subseteq Str, c \in C_i, c' \notin C_i\}$ of the subsystems. Hence, $S_i = (C_i, M, Id, Sd_i, Ctx, Obs_i, In_i, Out_i)$ is a subsystem to be diagnosed. A candidate diagnosis of the subsystem S_i is denoted by D_i.

Definition 7 *Let $S_i = (C_i, M, Id, Sd_i, Ctx, Obs_i, In_i, Out_i)$ is a subsystem to be diagnosed. Let $Obs_{\text{con}}, Obs_{\text{abd}} \subseteq Obs_i$ be two subsets of the observations, and let V_i be a (partial) descriptions of the values of the connection points In_i . Finally, let D_i be a candidate diagnosis. Then D_i is a diagnosis for S_i iff*

$D_i \cup Sd_i \cup Ctx \cup V_i \mid\sim \bigwedge_{\varphi \in Obs_{\text{abd}}} \varphi$ and $D_i \cup Sd_i \cup Ctx \cup V_i \cup Obs_{\text{con}} \not\mid\sim \bot$.

Given multiple diagnostic agents, an important question is how the diagnoses of the agents relate to the diagnoses of a single agent that has complete knowledge of the system description and the observations. When addressing this question we assume throughout the paper that *there are no conflicts between the knowledge of the different agents*. That is, there is a diagnosis D such that: $D \cup Sd \cup Cxt \cup Obs$ is consistent.

Proposition 4 *Let A_1 and A_2 be two diagnostic agents each having partial knowledge of the system; i.e. S_1 and S_2. Moreover, let D be a single agent diagnosis of S.*

Then for some (partial) descriptions V_1 and V_2 of the values of the connection points In_1 respectively In_2, $D_1 = \{mode(c, s) \mid c \in C_1, mode(c, s) \in D\}$ is a diagnosis of S_1 and $D_2 = \{mode(c, s) \mid c \in C_2, mode(c, s) \in D\}$ is a diagnosis of S_2.

[6] That is, D is a partial diagnosis if there are no *mode* conflicts; i.e. for no $c \in C$: $mode(c, nor)$, $mode(c, ab) \in D$.

Proposition 5 *Let A_1 and A_2 be diagnostic agents with partial knowledge S_1 respectively S_2. Moreover, let D_i be a diagnosis of S_i determined by agent A_i given some partial description V_i of the values of In_i .*

Then, $D = D_1 \cup D_2$ is a single-agent diagnosis if D is a candidate diagnosis and if $D_i \cup Sd_i \cup Ctx \cup V_i \mid\sim \bigwedge_{\varphi \in V_j} \varphi$ for every $j \neq i$.

Note that the above propositions show that multi-agent diagnosis is possible. In particular, Proposition 5 offers the possibility to establish global diagnoses by information exchange between agents

The complexity of determining a global diagnosis depends on the organization of the multi-agent system. First, knowledge of the system can be distributed in different ways over the agents. We will consider two extreme cases, knowledge that is either semantically of spatially distributed. Second, it makes an important difference whether agents use fault models of the behavior of components. Third, the dependencies between the knowledge distributions plays an important role. The dependencies determine whether agents have to exchange information to make a 'local' diagnosis.

Analysis

Dependent descriptions Before agents can establish a global diagnosis they first have to establish a local diagnosis using the knowledge of their part of the system. An important issue is whether they can do this independently of each other.

Dependencies arise because different models of the system are interconnected. By definition, such connections are present when knowledge is spatially distributed. When knowledge is semantically distributed, independence is possible, e.g., if an electrical and a thermodynamical description of the system is used. If, however, the heat of a (broken) component influences the electrical characteristics of the nearby components, we no longer have independence.

We can enforce independence by observing the values of all connection points between different descriptions of the system; i.e. the values of In_i. In large systems this may not be feasible. Hence, agents have to exchange predicted values of connection points for every candidate diagnosis they consider. This may cause large communication overhead since the number of candidate diagnoses is exponential.

Another problem is that the connection between subsystems may form *cycles*. Hence, predicting the behaviour of the system given a candidate diagnosis may require going through many cycles, causing large communication overhead.

Theorem 1 *Given a global candidate diagnosis D, predicting the values of all connection point is an NP-Hard problem.*

Semantically distributed knowledge If knowledge is semantically distributed, each agents looks at different aspects of the whole system. We will first consider the situation in which agents have no fault models, and in which the knowledge of the agents is independent. The latter implies that either there are no connections, $In_i = \varnothing$, between the different descriptions of the system or all connection points of the connections between S_1 and S_2 are observed.

If we only apply consistency based diagnosis we can derive the following result.

Proposition 6 *Let the diagnostic agents A_1 and A_2 be organized as described above and let D_1, D_2 respectively their diagnoses. Then, $D = \{mode(c, nor) \mid mode(c, nor) \in D_1, mode(c, nor) \in D_2\} \cup \{mode(c, ab) \mid mode(c, ab) \in D_1$ or $mode(c, ab) \in D_2\}$ is a single agent diagnosis.*

Note that if both D_1 and D_2 are minimal diagnoses, D need not be a minimal diagnosis.

As in the single agent approach, we can improve consistency based diagnosis if we also allow for abductive explanation of normal observations [8]. The results of Propositions 2 and 3 can be extended to multi-agent diagnosis.

Proposition 7 *Let the diagnostic agents A_1 and A_2 be organized as described above, let D_1^{abd} and D_2^{abd} be their abductive kernel diagnoses and D_1^{con} and D_2^{con} their consistency based kernel diagnoses. Then, $D = D_1^{abd} \cup D_2^{abd} \cup D_1^{con} \cup D_2^{con}$ is a single-agent kernel diagnosis if D is a partial diagnosis.*

Note that if both D_1 and D_2 are kernel diagnoses, D need not be a kernel diagnosis.

Proposition 4 implies together with propositions 6 respectively 7 that all minimal / kernel diagnoses can be determined in polynomial time if the number of minimal / kernel diagnoses of each subsystem is polynomial bounded in $|C|$.

Proposition 8 *There exists a* protocol[7] *that determines all global minimal / kernel diagnoses with a time complexity that is quadratic in the maximum of the number of minimal / kernel diagnoses of a subsystem.*

In some areas, it is important to know the type of fault that has occurred. In medical diagnosis for instance, we do not only need to know the component that is failing but also what is causing it to fail. In this area we usually do not replace a component but instead try to eliminate the cause of the malfunction. Hence, fault models are required.

Allowing for fault models complicates the process of combining the candidate diagnoses of several agents. The reason for this is that, given an ordering of candidate diagnoses $D_1 \prec D_2 \prec D_3 \prec D_4$, D_1 and D_3 can be diagnoses while D_2 and D_4 are not. Hence, we can no longer characterize an exponential number of diagnoses using a polynomial number of minimal or kernel diagnoses. Exchanging all (kernel) diagnoses between the agents is, in general, infeasible.

Proposition 9 *There exists a protocol that determines the* numerical *minimal diagnoses in $\mathcal{O}(t) \leq \mathcal{O}((n \cdot m)^k)$ time where $n = |C|$, $m = \max_{c \in C} |M_c|$, k is the number of broken components in a numerical minimal diagnosis and t is the number of local diagnoses with no more than k broken components of some subsystem.*

Spatially distributed knowledge If agents use fault models, they have to exchange information about the values of connection points in between subsystems for every candidate diagnosis they consider. The agents can reduce the amount of information exchange by ignoring the fault models. Agents may reduce the amount of information exchange even further if they may assume default values for these connection points. In both cases,

[7] Due to space limitations, we cannot present a protocol here.

we can only apply consistency based diagnosis or consistency based diagnosis with abductive explanation of normal observations.

Inputs of an agent's part of the system that are determined by other agents, can be incorrect. Therefore, agents must assume the correctness of these inputs and must be able to withdraw these assumptions during diagnostic reasoning. When an agent no longer assumes that an input is correct, it must pass on this information to the agent whose part of the system determines the input. For every candidate diagnosis an agent considers, it must provide this kind of feedback to the other agent(s). Since connections between subsystem may form loops, finding a minimal diagnosis can be hard.

To see how difficult finding a minimal diagnosis may be, view each subsystem as a variable and each candidate diagnosis of a subsystem as a domain value of this variable. Then finding a minimal diagnosis can be seen as a Constraint Optimization Problem.

Theorem 2 *Even if the agents have a polynomial algorithm for determining a local minimal diagnosis, determining a global minimal diagnosis is still an* NP-Hard *problem.*

5 Conclusion

Multi-agent diagnosis is possible but not always feasible. If diagnostic knowledge is semantically distributed, the usage of fault models may result in exchanging an exponential amount of information in order to establish a global diagnosis. If, however, the number of broken components is limited, then there exists a protocol that determines a global minimal / kernel diagnosis in polynomial time.

If diagnostic knowledge is spatially distributed, the amount of information exchange depends on whether the agents exchange predicted values. Circular dependencies between the information required by different agents may cause a lot of information exchange. Moreover, even without fault models, finding a global minimal diagnosis is an NP-Hard problem. Hence, future research should focus on efficient approximation protocols.

References

[1] L. Console and P. Torasso. Hypothetical reasoning in causal models. *International Journal of Intelligence Systems*, 5:83–124, 1990.

[2] L. Console and P. Torasso. A spectrum of logical definitions of model-based diagnosis. *Computational Intelligence*, 7:133–141, 1991.

[3] J. J. van Dixhoorn. Bond graphs and the challenge of a unified modelling theory of physical systems. In F. E. Cellier, editor, *Progress in Modelling & Simulation*, pages 207–245. Academic Press, 1982.

[4] P. Frohlich, I. de Almeida Mora, W. Nejdl, and M. Schroeder. Diagnostic agents for distributed systems. In J.-J. Ch. Meyer and P.-Y. Schobbens, editors, *Formal Models of Agents. ESPRIT Project ModelAge Final Report Selected Papers. LNAI 1760*, pages 173–186. Springer-Verlag, 2000.

[5] J. de Kleer, A.K. Mackworth, and R. Reiter. Characterizing diagnoses and systems. *Artificial Intelligence*, 56:197–222, 1992.

[6] J. de Kleer and B. C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32:97–130, 1987.

[7] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95, 1987.

[8] N. Roos. Efficient model-based diagnosis. *Intelligent System Engineering*, pages 107–118, 1993.

[9] M Schroeder and G. Wagner. Distributed diagnosis by vivid agents. In *Proceedings of the first conference on Autonomous Agents*, pages 268–275, 1997.