An analysis of Multi-Agent Diagnosis

Nico Roos
Universiteit Maastricht,
Infonomics / IKAT,
P.O.Box 616,
6200 MD Maastricht.

Annette ten Teije
Utrecht University,
ICS,
P.O.Box 80.089,
3508 TB Utrecht.

André Bos & Cees Witteveen
Delft University of Technology,
ITS,
P.O.Box 256,
2600 AJ Delft.

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 establish a diagnosis and 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 these agents?*'

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent systems

General Terms

Theory

1. INTRODUCTION

A traditional model-based 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 not be 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 is able to (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. [3]): (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 [6] in order to gain fault-tolerant behavior.

The way the knowledge is distributed turns out to have significant repercussions on multi-agent diagnosis.[1] We will show that, though multi-agent diagnosis turns out to be possible in theory, it is not always feasible.

2. THE DIAGNOSTIC SETTING

The global 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, the context 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.

AAMAS'02, July 15-19, 2002, Bologna, Italy.

---

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

The system description $Sd = Str \cup Beh$ consists of a structural description Str and a behavioral description Beh. The structural description Str describes the connections between components using the connection point identifiers Id. The behavioral description Beh specifies for each component $c \in C$ and for each (fault) mode in M_c of a component c, possibly with the exception of $ab \in M_c$, a behavior 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 $m \in M_c$.

A *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 $mode(c, m) \in D$.

The well-known concept of Model-Based Diagnosis [2] will be called *single agent diagnosis* since it assumes that a single agent, having complete knowledge of the system, S, suffices to make a diagnosis.

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

D is a diagnosis for S iff (1) $D \cup Sd \cup Ctx \mid\!\sim \bigwedge_{\varphi \in Obs_{\mathrm{abd}}} \varphi$, *(2)* $D \cup Sd \cup Ctx \cup Obs_{\mathrm{con}} \not\mid\!\sim \bot$. [3]

If $Obs_{\mathrm{abd}} = \varnothing$ and $Obs_{\mathrm{con}} = Obs$, then we have a pure *consistency-based diagnosis* [4, 5], and if $Obs_{\mathrm{con}} = \varnothing$ and $Obs_{abd} = Obs$, we have a pure *abductive diagnosis* [1].

In the Multi-Agent setting, we focus on knowledge of a system S that is *semantically* or *spatially* distributed. The distribution of the knowledge over the agents defines a division of the system into subsystems.

If knowledge is spatially distributed, the set of components C is partitioned over the agents. So, agent A_i has knowledge about components C_i, and $C = \biguplus_{i=1}^{m} C_i$ where m is the number of agents. If knowledge is semantically distributed, each agent possesses a different type of knowledge of the whole system. These knowledge types can be distinguished by the value types of the connection points. Ideally, the knowledge types are completely independent. This means that the behaviour of a component with respect to one knowledge type depends only on the mode of a component. For instance, the mechanical properties of a component are usually independent of its electrical properties. This may, however, not always be the case. For instance, power lines with high currents that close together, and electrical engines.

By distributing knowledge over the agents, we loose the knowledge about the connections between components managed by different agents. To compensate for this lack of information, we must provide each agent with information about the connection points that connect to components managed by other agents. To this end, we provide each agent with information about these connection points, and divide the connection points into relative inputs In_i and outputs Out_i of the agent's subsystem. An agent's subsystem is a tuple

$$S_i = (C_i, M, Id, Sd_i, Ctx, Obs_i, In_i, Out_i)$$

3. ANALYSIS

Each agent A_i must make a diagnosis of the subsystem S_i under its control. In order to this, the agent must know the values V_i of the inputs In_i of its subsystem S_i that are determined by the outputs of other subsystems S_j. This information extends the context of S_i.

The following two propositions show that multi-agent diagnosis is possible.

Proposition 1. *Given a diagnosis D of the whole system S, there always exists for each subsystem S_i a corresponding local diagnosis D_i given the input values V_i. Moreover, values V_i follow from the diagnoses D_j that determine In_i.*

Proposition 2. *Given a local diagnosis D_i of each subsystem S_i, there is a corresponding diagnosis D of S provided that the subsystems agree on the values V_i of the connection points In_i between the subsystems.*

In order to make a diagnosis, the agents must first predict the system's behavior.

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

Analysis of the sources of complexity of the problem shows that (1) we should either observe connection points between subsystems, or (2) we should avoid circular dependencies between subsystems, or (3) we should minimize the number of diagnoses that the agents must consider.

Determining a global diagnoses from local diagnoses can be done efficiently if knowledge is semantically distributed.

Proposition 3. *Let the knowledge be semantically distributed over the agents. Then there exists a protocol that enable the agents to determine all numerical minimum diagnosis D of S in polynomial time.*

Theorem 1 implies that in case of spatially distributed knowledge, the agents should restrict themselves to consistency based diagnosis. In case of consistency-based diagnosis, the agents predict the system's behavior for only one candidate diagnosis: no broken components. Unfortunately, this does not solve all problems.

Theorem 2. *Let knowledge be spatially distributed over the agents and let the agents only use consistency-based diagnosis. Then even if the agents have a polynomial algorithm for determining all local minimal diagnosis D_i, determining a minimal diagnosis D of S is still an* NP-Hard *problem.*

4. REFERENCES

[1] L. Console and P. Torasso. Hypothetical reasoning in causal models. *Int. 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] P. Frohlich, I. de Almeida Mora, W. Nejdl, and M. Schroeder. Diagnostic agents for distributed systems. In J.-J. C. Meyer and P.-Y. Schobbens, editors, *Formal Models of Agents. LNAI 1760*, pages 173–186. Springer-Verlag, 2000.

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

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

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

---

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

[3] The symbol $\mid\!\sim$ denotes the possibly limited reasoning capabilities of a diagnostic agent: $\{\varphi \mid \Sigma \mid\!\sim \varphi\} \subseteq \{\varphi \mid \Sigma \vdash \varphi\}$.