Reaching diagnostic agreement in Multi-Agent Diagnosis

Nico Roos
IKAT
Universiteit Maastricht
P.O.Box 616
6200 MD Maastricht
roos@cs.unimaas.nl

Annette ten Teije
Faculty of Sciences
Free University
De Boelelaan 1081A
1081 HV Amsterdam
annette@cs.vu.nl

Cees Witteveen
Faculty of EEMCS
Delft University of Technology
PO Box 5031
2600 GA Delft
c.witteveen@ewi.tudelft.nl

Abstract

We consider the problem of finding a commonly agreed upon diagnosis for errors observed in a system monitored by a number of different agents. Each agent is assumed to have its own specialized (expert) view on the system. Collectively, the agents have to agree on one or more diagnoses based on their views. Reaching an agreement is complicated by the fact that the knowledge of different specialists need not always be correct.

1. Introduction

A traditional diagnostic tool can be viewed as a single *diagnostic agent* having a model of the whole system to be diagnosed. In some applications, however, such a single agent approach is infeasible or at least undesirable. The introduction of expert agents immediately raises the problem how to reach an agreement on the *cause* of observed anomalies. As was pointed out in [3], assuming a fixed maximum number of broken components, there exists a polynomial time protocol for reaching an agreement between the agents in case of a such semantic knowledge distribution (cf. [2]).[1] This protocol is rather straight forward. A more difficult situation, however, arises if the knowledge of some agents is *incorrect*, i.e., the agents have incompatible knowledge about the behaviors of components. In this paper, we will address this issue.

2. The diagnostic setting

We have a system S to be diagnosed by a set of agents A. Knowledge of the system S is distributed over the agents A resulting in a subsystem $S_i = (C, M, Sd_i, Ctx, Obs_i)$ to be diagnosed the agent $i \in A$. Here, C is a set of components, $M = \{M_c \mid c \in C\}$ is a specification of behavior modes per component , Sd_i is the system description, Ctx is a specification of input values of the system and Obs_i is a set of observations made by agent i. C, M and Ctx are the same for all agent in A. 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. The system description Sd_i consists of a structural description Str known by all agents and a behavioral description Beh_i of the components from the perspective of agent i. The set Beh_i specifies a behavioral knowledge for each component $c \in C$. The behavioral knowledge of agent i about component c specifies the component's behavior for each mode $m \in M_c$ as an implication of the form $mode(c, m) \rightarrow \Phi_i$. With the exception of the mode ab, a behavior is specified for for every mode. A *candidate diagnosis* is a set D_i 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 $mode(c, m) \in D_i$.

Definition 1 *Let $S_i = (C, M, Sd_i, Ctx, Obs_i)$ be a subsystem from the perspective of agent i. Let $Obs_i^{\rm con}$, $Obs_i^{\rm abd} \subseteq Obs_i$ be subsets of the observations. Finally, let D_i be a candidate diagnosis of S_i.*
D_i is a diagnosis of S_i iff [1]:
$D_i \cup Sd_i \cup Ctx \vdash Obs_i^{\rm abd}$,
$D_i \cup Sd_i \cup Ctx \cup Obs_i^{\rm con} \not\vdash \perp$.

3. Agents with incorrect knowledge

Knowledge of agents about the components' behaviors may in some situation be incorrect. As a result, agents need not agree on the components that can be broken or on the fault modes of the broken components. A robust multi-agent system should be able to handle such situations. We use an *abstraction hierarchy* on the fault modes to deal with this problem. Here, agents have less detailed knowledge about less specific behavior modes. Since no behavior is specified for the mode ab, agents must be able to agree on the least

1 Besides a semantic knowledge distributed, we also distinguish a *spatial knowledge distribution* The latter has been discussed in [4].

specific diagnoses. Note that abductive diagnosis is not applicable for less specific fault modes.

A protocol for diagnostic agreement Since a (local) diagnosis of a specific agent need not be compatible with a common diagnosis, the agent proposing the diagnosis needs to receive feedback when its diagnosis is rejected by other agents. Subsequently, the agent can generate a new diagnosis taking into account the diagnoses that have been rejected. The generation of new diagnoses can be improved if other agents provide the reasons for rejecting a proposed diagnosis. (See the full paper for the complete protocol.)

4. Probable diagnoses

The protocol for diagnostic agreement determines a less specific diagnosis in case the agents cannot agree on a most specific diagnosis. In this way agents can reach an agreement even if the knowledge of some agent predicts the wrong behavior given the current context and current observations. What we would like to know is how this affects the probability of a diagnosis, especially if the knowledge of some agents is incorrect given the current context and current observations. In the full paper, probabilistic correctness measures for diagnoses have been derived for the case that the agents' knowledge is correct and the case that the agents' knowledge is not always correct. These measure enable the agents to determine the probability of each diagnosis. Note that the agents should first determine the components that are broken; i.e. the least specific diagnoses according to the abstraction hierarchy on the behavior modes. Subsequently, the agents may select more specific diagnoses, which may be less probable.

5. Experiments

We investigated whether the correct diagnosis is among the most probable diagnoses if the knowledge of one of the agents contains an error. In the experiments, we generated 8000 systems each to be diagnosed by three agents. We chose three agents since this is the smallest number to make one diagnosis significantly more probable if one of the agents disagree with the others, while using more agents would have simplified the diagnostic problem. Each generated system consisted of 40 components, each with one output and two inputs. An input was either connected to one of the four system inputs or to an output of a randomly chosen component without causing cycles.

The normal behavior of a component was a modulo n adder for each of the three agents each using a different perspective. Besides, a component had faulty behaviors, namely ab and two specific faulty behaviors f_1 and f_2. In both fault modes f_1 and f_2, a fault value was added modulo n to the output of the component. These faults values were randomly chosen for each combination of a component, a fault mode and an agent. Finally, for every component c, the same value was used for the probabilities of the fault modes f_1 and f_2 and the probability that a behavior mode is incorrect.

To create a diagnostic problem, in each generated system one component was chosen to be the broken component and one of the fault mode f_1 or f_2 was selected for the component. In one of the three perspectives, however, the component behaved according to the other fault mode, i.e. the knowledge of the agent using this perspective was incorrect in the current situation.

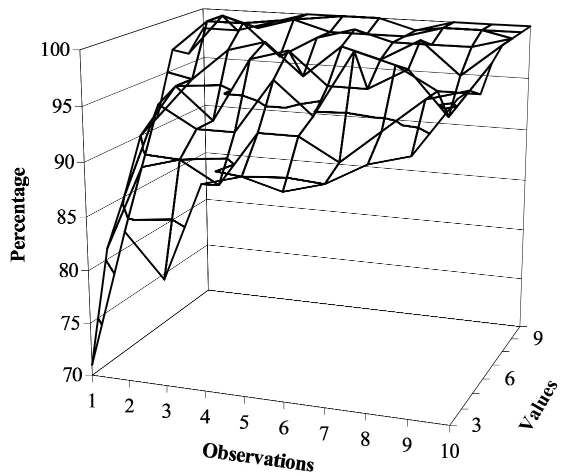

Figure 1. Distinguishability.

To create the most difficult problems, the agents all used the same -randomly chosen- observation points. Figure 1 shows the percentage of problems in which the correct diagnosis is among the most probable diagnoses. Moreover, the average number of probable diagnoses that were determined by the agents was 1.47.

References

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

[2] 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. ESPRIT Project ModelAge Final Report Selected Papers. LNAI 1760*, pages 173–186. Springer-Verlag, 2000.

[3] N. Roos, A. ten Teije, A. Bos, and C. Witteveen. An analysis of multi-agent diagnosis. In *AAMAS 2002*, pages 986–987, 2002.

[4] N. Roos, A. ten Teije, and C. Witteveen. A protocol for multi-agent diagnosis with spatially distributed knowledge. In *AAMAS 2003*, pages 655–661, 2003.