Model-based Diagnosis of Multi-Agent Systems

Meir Kalech
Advisor: Gal A. Kaminka
Bar Ilan University, Israel

Background and Motivation

With increasing deployment of distributed applications in complex, dynamic settings, there is an increasing need to also be able to respond to failures that occur in their coordinated operation, in order to facilitate recovery and reestablish collaboration. I refer to this type of diagnosis as "social diagnosis", since it focuses on finding causes for failures to maintain designer-specified relationships between sub-systems. Unfortunately, while the problem of detection has been addressed in the literature (Kaminka & Tambe 2000), social diagnosis remains an open question. Naive implementations of social diagnosis processes can require significant computation and communications overhead, which prohibit them from being effective as the number of sub-systems is scaled up, or the number of failures to diagnose increases. I thus seek to examine in depth the communication and computation overhead of diagnosis.

For instance, suppose a company has multiple sub-systems that must interact with each other to achieve a common task, e.g., two cellular phone stations (cells) that must pass a cellular phone call from one to another as the user moves. The two stations must coordinate on the frequencies involved and other salient information. Sometimes, due to failures in the transmitter or in the receiver, or due to intermittent faults in the system, this coordination fails, and the two stations remain un-coordinated. Discovering the problem can fairly easy (the call gets disconnected), but determining the cause for it (the abnormal component) can be very difficult.

I choose to use a model-based diagnosis (*MBD*) approach that (Reiter 1987; de Kleer & Williams 1987) relies on a model of the diagnosed system, which simulates the behavior of the system given the operational context (typically, the system's inputs). The resulting simulated behavior (typically, outputs) are compared to the actual behavior to detect discrepancies indicating failures. The model can then be used to pinpoint possible failing components within the system.

MBD is increasingly being applied in distributed and multi-agent systems (Roos, Teije, & Witteveen 2003; Lamperti & Zanella 2003). While successfully addressing key challenges, MBD has been difficult to apply to diagnosing coordination failures. This is because many such failures take place at the boundaries between the agent and its environment, including other agents. For instance, in the former example, a cell may pass a call to another cell that due to a broken receiver, did not receive it. As a result, the two cells come to disagree on an action to be taken.

In my research I present a novel approach to detect and diagnose coordination failures, given the actions of agents, and the coordination constraints that should ideally hold between them. In the example above, knowing that the two cells should be in agreement as to their actions, and seeing that their actions are not in agreement, I could (1) show that a coordination failure has occurred; and (2) to propose several possible diagnoses for it (e.g., the first cell did not pass the call, the second cell did not receive it, etc.).

I adopt a consistency-based approach, which enables to use a general reliable and robust diagnosis method which I believe is applicable to many multi-agent systems. In addition, this approach guarantees minimal diagnosis and concerns about minimal communications in disambiguating the correct diagnosis.

Disagreement Between Agents

In the first stage of my thesis I focused on the diagnosis of disagreement between behavior-based agents (Kalech & Kaminka 2003). I distinguished two phases of diagnosis: (i) selection of the diagnosing agents; and (ii) diagnosis of the global team state (by the selected agents). I provided alternative algorithms for these phases, and combined them in different ways, to present diagnosis methods, corresponding to different design decisions.

I empirically evaluated the communication and run-time requirements of these methods in diagnosing thousands of systematically-generated failure cases, occurring in a complex simulation application. The results showed that centralizing the disambiguation process is a key factor in dramatically improving communications efficiency, but is not a determining factor in run-time efficiency. On the other hand, explicit reasoning about the different sub-systems is a key factor in determining run-time: Methods that require explicit reasoning about different sub-systems incur significant computational costs, though they are sometimes able to reduce the amount of communications.

Based on the conclusion that centralizing the diagnosis reduces the communication, I addressed two principles to reduce the communication and the computation in teams where the number of agents is scaled up (Kalech & Kaminka 2005a). First, I disambiguated the information sent by the agent before communicating: Instead of sending all the information, send only the information that is relevant to the diagnosis. Second, I diagnosed a limited number of agents that represent all others. These principles yield a novel diagnosis method which identifies the minimal number of agents that are necessary in order to make the diagnosis process. This method significantly reduces the runtime, while keeping communications overhead to a minimum.

The Model and the Diagnosis Process

In the second stage of my thesis I generalized the diagnosis method to any relation between agents and formalized it using a model based diagnosis approach (Kalech & Kaminka 2005b). The multi-agent systems of interest to me are composed of several agents, which (by design) are to satisfy certain coordination constraints.

Two agents may be coordinated in regard to certain components. I utilized two coordination primitives—*concurrence* and *mutual exclusion*—to define the coordination constraints. Concurrence states that two specific actions must be taken jointly, at the same time. Mutual exclusion states the opposite, i.e., that two specific actions may not be taken at the same time. I modelled these coordination using logical statements. I called to the model of the agents plus the model of the coordination *a team model*.

A fault in the coordination of a multi-agent system may be the result of a fault in one of the components or other agent components (it may also be the result of a fault in the environment, e.g., when a message is lost in transit). Given a team model and a partial observation of the agents' components, the goal of the social diagnosis is to determine a minimal set of abnormal components of agents whose selection may explain the inconsistency of the system.

The diagnosing process takes in three steps:

1.Observation - the diagnoser observes the value of some of the components of the agents.

2. Coordination diagnosis - then it checks whether the model of the coordination of the agents is consistent with the observed components.

3. Reasoning - the diagnoser continues in a back-chaining process in order to disambiguate exactly the abnormal components which cause the candidate abnormal component.

A possible method to compute the diagnosis is to calculate all the assignments for the agents' variables that satisfy the coordination constraints between the agents. Then the diagnoser compares the existing assignments of the agents to those that will satisfy the coordination, and computes a *minimal* set of changes.

Up to here I described a centralized diagnosis process in which only one diagnoser observes and diagnoses the team. However, centralized diagnosis methods suffer from some drawbacks. First, they can be computationally expensive in practice, in terms of communications and run-time. Second, they rely on a single diagnoser, and thus risk a single point of failure. Moreover, they assume no communication limitations, e.g., range. In the last stage of my thesis I present a distributed approach using algorithms from distributed CSP discipline. In order to compute a minimal diagnosis the agents must finding the whole assignments for the agents' variables that satisfy the coordination constraints between the agents. However, even distributed approach involves a high number of communications and its runtime grows exponentially. Therefore, I presented distributed algorithms that compute only a single satisfaction. Obviously, such methods do not guarantee minimal diagnosis, and they do not compute the whole diagnosis space, but on the other hand, these methods are much more efficient and save up communications.

I evaluated the use of such algorithms in comprehensive experiments with a team of physical and simulated Sony Aibo robots, experiencing systematic coordination failures. I examined the computational requirements of the algorithms (i.e., their run-time and bandwidth usage), and the correctness of the diagnoses produced. I find that in general, synchronous backtracking methods that compute the entire space of minimal diagnoses are naturally more expensive than others, though they produced better diagnosis results.

References

de Kleer, J., and Williams, B. C. 1987. Diagnosing multiple faults. *Artificial Intelligence* 32(1):97–130.

Kalech, M., and Kaminka, G. A. 2003. On the design of social diagnosis algorithms for multi-agent teams. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-03)*, 370–375.

Kalech, M., and Kaminka, G. A. 2005a. Diagnosing a team of agents: Scaling-up. In *Proceedings of Autonomous Agents and Multi Agent Systems (AAMAS-05)*.

Kalech, M., and Kaminka, G. A. 2005b. Towards model-based diagnosis of coordination failures. In *American Association for Artificial Intelligence (AAAI-05)*.

Kaminka, G. A., and Tambe, M. 2000. Robust multi-agent teams via socially-attentive monitoring. *Journal of Artificial Intelligence Research* 12:105–147.

Lamperti, G., and Zanella, M. 2003. *Diagnosis of Active Systems*. Kluwer Academic Publishers.

Reiter, R. 1987. A theory of diagnosis from first principles. *Artificial Intelligence* 32(1):57–96.

Roos, N.; Teije, A. t.; and Witteveen, C. 2003. A protocol for multi-agent diagnosis with spatially distributed knowledge. In *Proceedings of Autonomous Agents and Multi Agent Systems (AAMAS-03)*, 655–661.