Model-based Diagnosis of Multi-agent Systems

Meir Kalech and Gal A. Kaminka
The MAVERICK Group
Computer Science Department
Bar Ilan University, Israel
{kalechm, galk}@cs.biu.ac.il

1. BACKGROUND AND MOTIVATION

One of the key requirements in collaborating distributed applications is that their sub-systems remain synchronized during their joint operation. There is an increasing need to be able to respond to failures that occur in their synchronization and coordinated operation, in particular to be able to diagnose the causes for synchronization failures that may occur.

We choose to use a model-based diagnosis approach (*MBD*) [1] that relies on a model of the diagnosed system, which is utilized to simulate 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. 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 their environment, including other agents.

For instance, in a team, an agent may send a message that another agent, due to a broken radio, did not receive. As a result, the two agents come to disagree on an action to be taken. Lacking an omniscient diagnoser that knows of the sending of the message, the receiver has no way to detect and diagnose its fault. Surprisingly, it is still often possible 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 agents should be in agreement as to their actions, and seeing that their actions are not in agreement, is sufficient to detect the fault and diagnose it.

We keep a model of both the agents as well as the coordination between the agents, which enables us to use a general reliable and robust diagnosis method which we believe is applicable to many multi-agent systems.

2. DISAGREEMENT BETWEEN AGENTS

In the first stage of my thesis I focused on the diagnosis of disagreement between behavior based agents. We distinguish two phases of diagnosis: (i) selection of the diagnosing agents; and (ii) diagnosis of the global team state (by the selected agents). We provide alternative algorithms for these phases, and empirically evaluated the communications and run-time. The results showed that centralizing the disambiguation process is a key factor in improving communications, but is not a determining factor in run-time. On the other hand, explicit reasoning about the different sub-systems is a key factor in determining run-time.

Based on this conclusion we addressed two principles to achieve the reduction of the communication and the computation in large-scale teams. First, instead of sending all the information, send only the information that is relevant to the diagnosis. Second, we diagnose a limited number of agents that represent all others. These principles yield a novel diagnosis method which significantly reduces the runtime, while keeping communications overhead to a minimum.

3. COORDINATION FAILURES

In the second stage of my thesis I generalized the diagnosis method to any relation between agents and I formalized it using model based diagnosis approach. The multi-agent systems of interest to us are composed of several agents, which (by design) are to satisfy certain coordination constraints. We utilize two coordination primitives: *concurrence* and *mutual exclusion*. Concurrence states that two specific actions must be taken jointly, at the same time. Mutual exclusion states the opposite. We model these coordination using logical statements.

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) the diagnoser observes the agents, then (2) it checks whether the model of the coordination of the agents is consistent with the observation. If not, it computes the minimal set of abnormal agents using MBD techniques. Finally (3) it continues in a back-chaining process in order to disambiguate exactly the abnormal components.

4. REFERENCES

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

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

[3] M. Kalech and G. A. Kaminka. Diagnosing a team of agents: Scaling-up. *In Proceedings of the International Workshop on Principles of Diagnosis (DX 2004)*, pages 129–134, 2004.

AAMAS'05, July 25-29, 2005, Utrecht, Netherlands.