Goal recognition, obfuscation-based decision making and cooperation

Keywords: Obfuscation-based decision making, Sequential plan recognition, Behavior analysis, Cooperation, Multiagent systems

Extended Abstract

Imagine a scenario where agents representing countries decide to build and use nuclear-related industrial facilities to reach goals such as producing energy, acquiring nuclear weapon or producing materials for medical purposes. As they are also observing each-other, they need to integrate in their decisions the information they provide to potential onlookers. They may willingly obfuscate their goals or, on the opposite, make them as clear as possible.

Obfuscation is neither conceptualized here as obscurantism or deception. Agents may both do obfuscation and willingly reveal their intended goals to other trusted agents, and in that way contirubte to the transparency of the system. Obfuscation is a manner to have more control and select who you want (or should) be transparent with. Obfuscation must also be distinguished from deception as it does not broadcast fake information nor delude observers, but focuses on (and minimizes) the information given to unwanted observers.

Observing and interpreting the behavior of the agents in a multiagent system has been often mentioned as a necessity to identify threats, misconducts or any improper goal and eventually react in an appropriate way. This task is considered as a fundamental problem in AI called the *Plan recognition problem* [3] with many practical applications as computer network security[1] or human needs recognition. Plan recognition algorithms have been proposed to get information and infer the most probable goal of the observed agent.

But in many application domains, especially where privacy and safety of the users are involved, being able to identify the goals of the others might be considered as intrusive and maleficent. As a result agents tend, in different contexts and for various reasons, to willingly ignore information about the behavior of others [2, 4].

The main contribution of this work lies in the reasoning about the impact of a decision in terms of given information to a potential observer. The information given to an observer is used as an element of the context in the decision making process of an agent.

To integrate this awareness in the decision making process, we propose a mechanism designed to provide an evaluation of the quantity of given information, as a new element of the context of a decision. To do so, the agent explores the set of all possible explanations of her behavior. An explanation is defined as a minimal forest of plan trees with pending sets recorded for each node sufficient to allow the assignment of each observation to a specific action in the plans. The most probable plan is the plan which is most associated with actions in the set of all possible explanations for a given behavior. An *obfuscation-based decision making agent* always takes decisions in order to maximize entropy in the estimated probabilities of plans, i.e. to keep the probability of the most probable plan in the eye of the observer as low as possible.

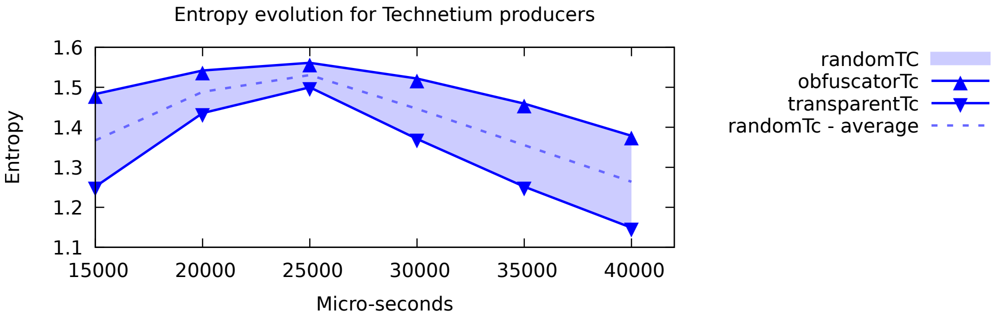

Figure 1: Observation of the evolution of the entropy in sequential plan recognition

We consider in this work the case where agents have to collaborate in a shared environment where they are able to observe, at least partially, the behavior of the others. Agents may willingly avoid to cooperate with those who do or don't obfuscate their own goals. This restriction makes sense for instance if agents care about privacy, or want to maximize uncertainty on the goal of the agents in the system for a strategic purpose.

A proof of concept has been implemented in a classic BDI agent architecture [5], using the JaCaMo Multiagent Systems framework based on AgentSpeak [6], in a realistic scenario based on the nuclear facilities construction example presented previously. Figure 1 illustrates the evolution of the uncertainty of an onlooker observing the behavior of a group of agents sharing the same goal, and shows how the obfuscation-based decision makers are always maximizing the uncertainty for the plan recognition process, whereas transparency-based decision makers are minimizing it, compared with all the possible behavior leading to achieve the same goal. In this example, obfuscators decide to build nuclear facilities which are less specific to technetium production. On the opposite, transparency-based decision makers decide to build the most specific type of material testing reactor to let their intended goal be as obvious as possible.

References

[1] Christopher W Geib and Robert P Goldman. Partial observability and probabilistic plan/goal recognition. In *Proceedings of the International workshop on modeling other agents from observations (MOO-05)*, volume 8, pages 1–6, 2005.

[2] Russell Golman, David Hagmann, and George Loewenstein. Information avoidance. *Journal of Economic Literature*, 55(1):96–135, 2017.

[3] Reuth Mirsky, Roni Stern, Kobi Gal, and Meir Kalech. Sequential plan recognition: An iterative approach to disambiguating between hypotheses. *Artificial Intelligence*, 260:51–73, 2018.

[4] Karine Nyborg. I don't want to hear about it: Rational ignorance among duty-oriented consumers. *Journal of Economic Behavior & Organization*, 79(3):263–274, 2011.

[5] Anand S Rao and Michael P Georgeff. Modeling rational agents within a BDI-architecture. pages 473–484, 1991.

[6] Renata Vieira, Álvaro F Moreira, Michael Wooldridge, and Rafael H Bordini. On the formal semantics of speech-act based communication in an agent-oriented programming language. *Journal of Artificial Intelligence Research*, 29:221–267, 2007.