Norm Monitoring with Asymmetric Information

(Extended Abstract)

Felipe Meneguzzi
School of Computer Science
PUCRS
Porto Alegre, RS 90619, Brazil
felipe.meneguzzi@pucrs.br

Brian Logan
School of Computer Science
University of Nottingham
Nottingham, NG8 1BB, UK
bsl@cs.nott.ac.uk

Moser Silva Fagundes
School of Computer Science
PUCRS
Porto Alegre, RS 90619, Brazil
moser.fagundes@pucrs.br

ABSTRACT

In this paper we consider the implications of imperfect monitoring in a stochastic environment for both the agents and the normative organisation in a normative MAS. We introduce a notion of *information asymmetry* to characterise the agents' knowledge of the monitoring strategy, and show that there are potential benefits of information asymmetry for the normative organisation in reducing its cost of enforcement.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*intelligent agents, multiagent systems*

General Terms

Algorithms, Design, Theory

Keywords

Norms, NMDPs

1. INTRODUCTION

Recent research on norm regulated behaviour in multiagent systems has resulted in a variety of approaches that allow individual agents to reason about norm compliance, particularly when compliance conflicts with the achievement of the agent's goals or the maximisation of utility. These approaches generally make three strong assumptions: (a) the environment is deterministic; (b) norm monitoring and enforcement are perfect; and (c) agents are fully aware of the monitoring capabilities of the normative organisation. For example, in much of the work on norm-aware agency, e.g., [3, 1], the agents implicitly assume that all norm violations will be detected and choose an 'optimal' course of action based on this assumption. However many environments are stochastic, and for large-scale, 'realistic' scenarios, perfect monitoring is likely to be either impossible or unfeasibly costly. In reality, the probability that violations of a norm will be detected (which we term the *enforcement intensity* of the norm) is likely to be less than 1. Moreover, in many multi-agent systems, it is reasonable to assume that complete information about the enforcement intensity employed by the normative organisation is not available to the agents at zero cost. In this case, agents must compute an optimal policy either by making an assumption about the enforcement intensity, or by learning.

In this paper we consider the implications of imperfect monitoring in a stochastic environment for both the agents and the normative organisation in a normative MAS. We show that if the agent makes an incorrect assumption about the enforcement intensity of a norm, its 'optimal' policy may not be optimal with respect to the norm, i.e., it could increase its utility by violating fewer norms or more norms, depending on whether the enforcement intensity is higher or lower than it assumes.

2. NORMATIVE MAS

Definition 1. A norm is a tuple $\langle \delta, \mathcal{G}, \phi, \psi, \rho \rangle$ where: $\delta \in \{$*obligation, prohibition*$\}$ is the deontic modality; $\mathcal{G}$ is the agent roles to which the norm applies; ϕ is the activation condition; ψ is the normative condition; and $\rho : \mathcal{S} \to \mathbb{R}$ is the penalty for violating this norm.

Determining when a norm is activated in a state, whether an activated norm is obeyed or violated, and (in the case of violations) for applying the appropriate penalty, is the responsibility of a *normative organisation.* Given a set of norms $\mathcal{N}$ and a set $\mathcal{S}$ of states of a normative MAS, a norm $n = \langle \delta, \mathcal{G}, \phi, \psi, \rho \rangle \in \mathcal{N}$ is violated in a state $s \in \mathcal{S}$ iff $\delta = \textit{obligation} \wedge s \models \phi \wedge \neg\psi$ or $\delta = \textit{prohibition} \wedge s \models \phi \wedge \psi$.

We assume that the probability that violations of a norm will be detected is under the control of the normative organisation. The *enforcement intensity* of the norm is a measure of the 'effort' the normative organisation is prepared to invest in detecting violations of the norm. An enforcement intensity of 1 indicates violations will be detected with cetainty, while and enforcement intensity of 0 indicates that the norm is not enforced (no violations are detected). The enforcement intensity is modelled as a detection function $\mathcal{D}(n)$, which gives the detection probability of the violation of the norm $n \in \mathcal{N}$ at any point in time.

Definition 2. A Normative Markov Decision Process (NMDP) is a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \mathcal{N}, \mathcal{D} \rangle$ where: $\mathcal{S}$ is the finite set of states of the world; $\mathcal{A}$ is the finite set of actions; $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ is the transition function where $\mathcal{T}(s, a, s')$ indicates the probability of executing a at s and ending at s'; $\mathcal{R} : \mathcal{S} \to \mathbb{R}$ is the reward function where $\mathcal{R}(s)$ is the utility gained by the agent for visiting the state s; $\mathcal{N}$ is the set of norms; and $\mathcal{D} : \mathcal{N} \to [0, 1]$ is the detection function,

Appears in: *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), Lomuscio, Scerri, Bazzan, Huhns (eds.), May, 5–9, 2014, Paris, France.*

which indicates the probability that a violation of norm n will be detected by the normative organisation.

With perfect information about $\mathcal{D}(n)$, an NMDP agent can construct a selfish utility-maximising policy using an approach similar to Fagundes et al. [2]. In this paper, we consider NMDP agents with incomplete information, i.e., the agent can only infer an approximation of the true detection probability $\tilde{\mathcal{D}}(n)$. This difference in information about norm enforcement leads to our main contribution, namely, that of *information asymmetry*.

3. INFORMATION ASYMMETRY

Definition 3. Let $\mathcal{N}$ be a set of norms enforced by a normative organisation with a detection function $\mathcal{D}(n)$ for all $n \in \mathcal{N}$, and let $\tilde{\mathcal{D}}(n)$ be the detection function known by an arbitrary agent in a multiagent system. We say that there is *information symmetry* in a normative MAS if $\forall n \in \mathcal{N}(\tilde{\mathcal{D}}(n) = \mathcal{D}(n))$. Conversely, we say that there is information asymmetry if $\exists n \in \mathcal{N}(\tilde{\mathcal{D}}(n) \neq \mathcal{D}(n))$

By varying the enforcement intensity $\mathcal{D}(n)$ for each agent when it joins a normative MAS, the normative organisation can bias the learning of an agent so as to induce the agent's behaviour. For example, the normative organisation may use a higher than usual enforcement intensity when the agent joins the MAS (and thus has a high α value), inducing an inflated value for $\tilde{\mathcal{D}}$ and then lower the value of $\mathcal{D}$ once the agent's update rule has converged. In this way, the normative organisation can exploit information asymmetry to reduce its cost of enforcement, by *inducing agents to behave as if the enforcement intensity is higher than it actually is.*

To illustrate how information asymmetry can be exploited by a normative organisation, we present a simple example scenario, the *Parking World*, in which an agent has limited information about the enforcement intensity of a norm.

The Parking World consists of a 5×5 grid of cells. Cell $(1, 1)$ is the start state, and cell $(5, 5)$ is the end state. The agent can move from cell to cell orthogonally and can also perform a null action (which leaves it in the same cell). In addition, the environment contains a 'no-parking cell' $(3, 3)$ in which stopping is prohibited. The agent receives a positive reward of 1 for reaching $(5, 5)$, and a small negative reward of -0.04 for visiting all cells other than the no-parking cell. If the agent stops in the no-parking cell and the violation of the norm is detected (i.e., the norm is enforced), the agent receives a sanction of -1. If the violation is not detected, the agent receives a positive reward of 0.5, i.e., violating the norm and parking illegally is beneficial.

To illustrate the effects of an agent's assumptions about the enforcement intensity on the rewards it obtains, we consider a simple scenario in which agents learn an optimal policy for ten instances of the Parking World where the enforcement intensity of the no-parking norm ranges from 0 to 1.0 in 0.1 steps. Agent 1 learns an optimal policy for an enforcement intensity of 0, Agent 2 learns an optimal policy for an enforcement intensity of 0.1, . . . , and Agent 11 learns an optimal policy for an enforcement intensity of 1.0. After learning, all the agents are placed in a Parking World where the enforcement intensity is 0.5. Figure 1 shows the difference between the reward each agent expects and the reward it actually receives ($\Delta U =$ actual reward$-$expected reward)

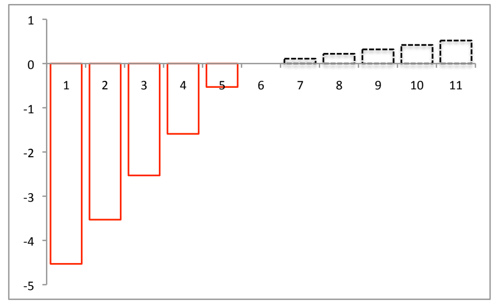

Figure 1: Actual minus expected reward under information asymmetry.

for the no-parking cell (3, 3) in the new environment. As can be seen, Agents 1–5 that under-estimate the enforcement intensity receive a lower reward than they expect as they incur a larger than expected number of sanctions. Conversely, Agents 7–11 over-estimate the actual enforcement intensity and could obtain a higher reward by exploiting the lower than expected enforcement of the parking norm, but if they act on their policy, they will obtain their expected reward.

4. DISCUSSION

Although simple, the Parking World serves to highlight the importance of information asymmetry. If the agent's policy under-estimates the true enforcement intensity, the agent receives a clear signal that its policy is incorrect in the form of (unexpected) sanctions and a lower than expected reward. However an agent with a fixed (or slowly changing) policy that over-estimates the enforcement intensity receives no such signal from the environment and thus has no reason to change its policy. It will continue to act on its policy believing it to be correct. In particular, its degree of compliance with the norm will be higher than an agent with perfect information. Even if the agent does continue to learn, the difference in utility is less pronounced when the agent over estimates the actual enforcement. This information asymmetry can be exploited by the normative organisation to reduce the cost of enforcement.

The notion of information asymmetry lays the foundation for a number of avenues of future work, which we have just started to explore. In particular, we plan to investigate the interplay between agent learning strategies and normative monitoring strategies.

5. REFERENCES

[1] N. Alechina, M. Dastani, and B. Logan. Programming norm-aware agents. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, pages 1057–1064, 2012.

[2] M. S. Fagundes, S. Ossowski, M. Luck, and S. Miles. Using normative markov decision processes for evaluating electronic contracts. *AI Communications*, 25(1):1–17, 2012.

[3] F. Meneguzzi and M. Luck. Norm-based behaviour modification in BDI agents. In *Proceedings of the Eighth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, pages 177–184, 2009.