

Reasoning About Norms Within Uncertain Environments

(Extended Abstract)

N. Criado
Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia
Camino de Vera, s/n. 46022. Valencia, Spain
ncriado@dsic.upv.es

ABSTRACT

The main aim of my thesis is the development of agents capable of reasoning about norms given that they are situated in an uncertain environment. The n-BDI agent architecture developed in my thesis is aimed at allowing agents to determine which and how norms will be obeyed and supporting agents when facing with norm violations.

Categories and Subject Descriptors

I.2.11 [Distributed Artificial Intelligence]: Intelligent agents

General Terms

Theory

Keywords

Norm compliance, BDI agents, Uncertainty

1. INTRODUCTION

Internet is, maybe, the most relevant scientific advance of our days. It has also allowed the evolution of traditional computational paradigms into the paradigm of distributed computation over a open network of machines [11]. Multi-agent systems (MAS) have been proposed as a suitable technology for addressing challenges motivated by these open distributed systems. MAS applications are formed by agents which may be designed independently according to different goals and motivations. Therefore, no assumption about their behaviours can be made *a priori*. Because of this, coordination and cooperation mechanisms, as *norms*, are needed in MAS for ensuring social order and avoiding conflicts [2].

In MAS research, norms have been defined as a formal specification of what is permitted, obliged and forbidden within a society. Thus, they aim at regulating the life of software agents and the interactions among them [12]. Norms have been proposed in MAS to deal with coordination issues [10], to model legal issues in electronic institutions and electronic commerce [8], to model MAS organizations [7].

Cite as: Reasoning About Norms Within Uncertain Environments (Extended Abstract), N. Criado, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. 1331-1332.

Copyright © 2011, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

2. MOTIVATION

In this section, I pose the main questions that my thesis tries to answer. Fundamentally, it has been motivated by the fact that existing proposals of intelligent norm-aware agents, like [9, 3], tend to be concerned about the decision-making processes that are supported by a set of active norms whose validity is taken for granted. Thus, they consider norms as static constraints that are hard-wired on agents. Only a fraction [1] have been concerned about the fact that norms can be violated deliberately and rationally. Thus, in my thesis I will address the problem of defining norm-aware agents and, in particular, I discuss how these agents deliberate about norms within uncertain environments. This question raises the matter of what means to reason about norms. The work of Sripada et al. [14] analyses the psychological reasoning subserving norms. This process is formed by two closely linked innate mechanisms: one responsible for the norm compliance dilemma, deciding whether one observes or violates a norm at a given moment; and the other in charge of norm implementation, which detects norm violations and generates motivations to punish norm violators. The first question addressed by my thesis is:

- How to built agents capable of facing with the norm compliance dilemma within uncertain environments?

Regarding the first issue, the norm compliance dilemma may be defined intuitively as making a choice between obeying or violating norms. The question implies the development of agents capable of considering norms. The set of norms which regulate MAS may dynamically evolve along time. Therefore, agents must be able to recognise and adopt new norms but maintaining their autonomy. Once an agent recognises a norm it may consider the effect of norm compliance in order to decide between norm violation or obedience. My thesis will consider also the “rational violation of norms” [4], which is an interesting issue that has not received enough attention in the existing literature. Therefore, my work will consider violations not as random or rebellious acts. On the contrary, the notion of rationality (which include both self-interest, emotional and cooperative motivations) as a criterion for making a choice between obeying or violating norms will be explored.

- How to built agents capable of implementing norms within uncertain environments?

On the other hand, this second question implies the consideration of the norm implementation within real scenarios. In this sense, traditional models of norm implementation

have been built assuming the existence of a shared reality which is *certainly* observed by agents. However, in real scenarios agents interact within an *uncertain* environment. In this sense, the uncertain environment implies a drastic evolution of the determination of norm violations. Up to the moment, sound norm violations have been detected by observing agent behaviour. Uncertainty about norm violation is explained by two main reasons: the opacity and limited knowledge about actions and illocutions performed by agents; and the existence of subjective conditions of norm violation due to the ambiguous interpretation of norms. Moreover, norm violations may be caused since agents are either unaware of the existence of the norm or do not perceive the discrepancy between the norm and their behaviour. Thus, norms imply processes for determining if a violation has occurred according to what has been observed by agents.

3. PHD THESIS APPROACH

In my thesis, my aim is to answer the question of the norm reasoning considering the inherent problematic of uncertain environments. As a response to this need, I will propose a normative BDI architecture (or n-BDI for short) [5, 6] in order to allow agents to take pragmatic autonomous decisions considering the existence of norms. Thus, the n-BDI will include an explicit representation of norms. These norms will allow normative desires and intentions to be inferred. Thus agents may exhibit both normative and non-compliant conduct. Rationality, emotionality and coherence will be the fundamental pillars of the n-BDI agent architecture. More concretely, rational motivations consider both: self-interest motivations, which consider the influence of norm compliance and violation on agent's goals; and the expectations of being rewarded or sanctioned by others. Non-Rational factors are related to internalised emotions such as honour and shame that maintain norms. Finally, coherence theory [15] will be employed as a criterion for determining which of these decisions are consistent with the current agent's mental state and how to build coherent alternatives for these decisions. In this sense, coherence among actions and goals will be considered in order to determine feasible plans for complying or violating norms.

Therefore, the combination of rationality, emotionality and coherence will allow agents to face the norm compliance dilemma in a more realistic way. Besides that, the normative reasoning not only implies making a decision about norm compliance but also being able to detect and react to violations committed by others. This is one of the main contributions of my thesis, the consideration of the detection, reacting and solving norm violations within uncertain environments. Uncertainty entails complex and significant difficulties which have not been considered by the previous proposals. These issues are related to the fact that there is not fully observability of the interaction performed by others. In addition, the way in which agents affect in the environment is imperfect. Thus, they may violate norms unconsciously. Finally, norms have not an unambiguous interpretation. Thus, violations are not detected by simply evaluating the truth value of logical formulas which represent norms. On the contrary, conflicts among agents about what is considered as an illicit act may arise. Thus, norms are not logic formulas but rather agreement processes for reaching a consensus about the occurrence of norm violations. This is an original perspective of the norm compliance problem

which has not been deeply considered before by works on the individual norm reasoning. In my opinion, this question is of outstanding importance for the success of agent-based software solutions for large-scale distributed problems. Therefore, my thesis I will also be focused on building agents endowed with capabilities for evaluating partners accordingly to norms from this complex and realistic perspective.

4. ACKNOWLEDGMENTS

This work was partially funded by the Spanish government under CONSOLIDER-INGENIO 2010 (CSD2007-00022) and by the FPU grant AP-2007-01256 awarded to N. Criado.

5. REFERENCES

- [1] G. Andrighetto, M. Campenní, F. Cecconi, and R. Conte. How agents find out norms: A simulation based model of norm innovation. In *NORMAS*, pages 16–30, 2008.
- [2] G. Boella, L. van der Torre, and H. Verhagen. Introduction to the special issue on normative multiagent systems. *Autonomous Agents and Multi-Agent Systems*, 17:1–10, 2008.
- [3] J. Broersen, M. Dastani, J. Hulstijn, Z. Huang, and L. van der Torre. The boid architecture – conflicts between beliefs, obligations, intentions and desires. In *AAMAS*, pages 9–16. ACM Press, 2001.
- [4] C. Castelfranchi. Formalising the informal? Dynamic social order, bottom-up social control, and spontaneous normative relations. *Journal of Applied Logic*, 1(1-2):47–92, 2003.
- [5] N. Criado, E. Argente, and V. Botti. A BDI Architecture for Normative Decision Making (Extended Abstract). In *AAMAS*, pages 1383–1384, 2010.
- [6] N. Criado, E. Argente, and V. Botti. Normative Deliberation in Graded BDI Agents. In *MATES*, volume 6251 of *LNAI*, pages 52–63. Springer, 2010.
- [7] V. Dignum, J. Vázquez-Salceda, and F. Dignum. OMNI: Introducing social structure, norms and ontologies into agent organizations. In *ProMAS*, volume 3346 of *LCNS*, pages 181–198. Springer, 2004.
- [8] A. García-Camino, P. Noriega, and J. A. Rodríguez-Aguilar. Implementing norms in electronic institutions. In *EUMAS*, pages 482–483, 2005.
- [9] M. Kollingbaum and T. Norman. NoA—a normative agent architecture. In *IJCAI*, volume 18, pages 1465–1466, 2003.
- [10] F. López y López, M. Luck, and M. d’Inverno. Constraining autonomy through norms. In *AAMAS*, pages 674–681, 2002.
- [11] M. Luck, P. McBurney, O. Shehory, and S. Willmott. Agent technology: Computing as interaction: A roadmap for agent-based computing. Technical report, Agentlink, 2005.
- [12] R. Rubino and G. Sartor. Preface. *Journal of Artificial Intelligence and Law*, 16(1):1–5, 2008.
- [13] C. Sripada and S. Stich. A framework for the psychology of norms. *The Innate Mind: Culture and Cognition*, pages 280–301, 2006.
- [14] P. Thagard. *Coherence in Thought and Action*. The MIT Press, Cambridge, Massachusetts, 2000.