Reputation-Based Decisions for Cognitive Agents (Thesis Abstract)

Isaac Pinyol
Artificial Intelligence Research Institute
Campus UAB, Bellaterra
Barcelona, Spain
ipinyol@iiia.csic.es

1. INTRODUCTION

Computational trust and reputation models have been recognized as key to design and implement multiagent systems [6]. These models manage and aggregate the information needed by agents to efficiently select partners in uncertain situations. In open multiagent systems agents have unknown intentions and thus, some kind of *interaction control* is necessary to ensure a well-fare society. Several approaches can be taken for this endeavor, each of them providing certain level of control. At the security level, the use of cryptography and digital signatures ensures privacy, integrity and authenticity of messages. At the organization level it is possible to define protocols and norms that agents must follow to interact, for instance, by defining electronic institutions. Finally, at the social level, reputation and trust models endow agents with a powerful social control artifact that permits them to *evaluate* potential partners considering certain criteria before the interaction is produced.

In recent years several reputation and trust models have been developed [10]. Most of them following game theoretical approaches prepared to deal with relatively simple environments. However, if we want to undertake problems found in socially complex virtual societies, more sophisticated trust and reputation systems based on solid cognitive theories are needed. One such cognitive theory is defined in [3].

This theory focuses on the impact that social evaluations have in the mental state of the agents and not on the computation of such evaluations. In this sense, the theory proposes that agents evaluate the performances of other agents according to certain criteria, and that these evaluations (social evaluations) can be only believed by the agents, only communicated by the agents or both believed and communicated. When a social evaluation is believed by a group of agents the theory refers to it as *image*. On the contrary, when a social evaluation circulates in the society it is refereed as *reputation*.

From this generic overview, the theory then develops a more individualistic vision. From a single agent, it describes a typology of possible decisions that autonomous agents can make involving social evaluations:

- *Epistemic decisions* cover the decisions about updating and generating social evaluations.
- *Pragmatic-strategic decisions* are decisions of how to behave with potential partners using social evaluations information, and thus, how agents use them to reason.
- *Memetic decisions* refer to the decisions of how and when to spread social evaluations.

Traditionally, the field of computational trust and reputation systems has been focused on developing and formalizing models as providers of social evaluations: on epistemic decisions. However, little attention has been paid to pragmatic-strategic and memetic decisions. This doctoral thesis embraces then these two *types* of decisions.

Currently, agents' decisions of how to use reputation information and how and when to spread it have been designed *ad-hoc* lacking any systematic or formal procedure. We claim that due to the cognitive nature of social evaluations, when facing complex societies pragmatic-strategic and memetic decisions can be as important as epistemic decisions. From this perspective, for a cognitive agent, the way a social evaluation is build can have the same importance as the final evaluation.

Under this scenario the thesis analyzes the integration of a particular cognitive reputation model, Repage [11], into a cognitive agent architecture, *Belief, Desire Intention* (BDI).

Taking Repage as a paradigmatic example of cognitive reputation model, the integration allows us on the one hand, to properly formalize the logical reasoning process of a cognitive agent where reputation information is implicitly taken into account. Therefore, we provides a formal framework that directly faces pragmatic-strategic decisions. On the other hand, the logical reasoning process can be seen as a way to build arguments over agents' attitudes [7], and these arguments can be used in negotiation processes, persuasion, information exchange or for simply explanatory purposes. Thereby, each action, intention, desire and belief of an agent can be justified by building an argument that can include of course reputation information. Thus,we are able to offer a formal framework in which memetic decisions are formalized in the context of argumentation frameworks.

2. OBJECTIVES AND DEVELOPMENT

In this section we detail the main objectives of the thesis.

2.1 Integration of Repage in a BDI Architecture

As we mentioned, Repage[11] is a computational reputation system based on a cognitive theory or reputation [3], and whose main characteristic is the distinction between image (what agents believe) and reputation (what agents said) in terms of social evaluations. Taking this model as example, we specify a BDI agent architecture as a multicontext system where Repage information is completely integrated. The reasoning process of the agent shows how pragmatic-strategic reputation-based decisions are taken in a formal and systematic way. The research done in this specific task is described below:

- **Definition of probabilistic dynamic belief logic** Since Repage uses probabilities and actions when describing social evaluations, we defined a belief logic capable to capture all the information that Repage provides [9].
- **Specification of the BDI agent:** To capture in a formal dimension the reasoning process of an agent, we specify a BDI architecture where its belief based is described using the formalism stated above. This required to use also graded desires and intentions for a correct integration. The underlying ideas for specifying this model were taken from the work by Casali and colleagues [2].
- **Description of study cases:** One of the most important task is to describes scenarios to enhance the relevance and potential of the model, demonstrating the advantages of paying attention to the integration models in the reasoning process. We provide simple scenarios where this necessity is proved.

2.2 Argumentation on Social Evaluations

Focusing on memetic decisions, we take advantage of the BDI+Repage defined above to build a generic argumentation framework where reputation information is also present in the arguments. As we mentioned in the introduction, argumentation can be used in different interaction processes, like negotiation protocols or even simple information exchange, to give more strength to the communicated information. In this point our work include:

- **Definition of a formal argumentation framework:** As we mentioned, we defined our BDI+Repage as a multicontext system [5]. Some work have shown how multicontext systems can be used to build argumentation frameworks[7]. Then, taking this approach we defined also Repage as a multicontext system to define an argumentation framework where each agent's attitude can be justified also with information from Repage. This implies for instance that certain intention can be supported by desires and beliefs, and that these beliefs can be also justified by the information coming from Repage.
- **Study Cases:** After the generic framework is defined, we apply it to concrete negotiation or information exchange protocols.

2.3 Implementation and Simulations

The theoretical work is complemented with implementation and simulation results to show the performances of our models facing concrete scenarios. Thus, our work incorporates empirical results to show how the theoretical aspects can be instantiated with current platforms. We focus on the following aspects:

- **Prototypical Implementation:** The BDI+Repage model is implemented using JASON [1], a multiagent platform that offers the advantages of logic programming together with functionalities to define multiagent scenarios. Of course, a direct implementation of our theoretical models is not feasible, due the computational complexity. However, with appropriated simplifications and assumptions instantiations are more than possible, and even capable to provide massive simulation results.
- **Verification using Simulation:** Using a BDI+Repage implementation we put the model to work by defining multiagent environments where cognitive agents have to deal with bad/good reputation information in competitive markets, following some previous work ([8],[4]).

3. REFERENCES

[1] R. H. Bordini, J. F. Hübner, and M. Wooldridge. *Programming Multi-Agent Systems in AgentSpeak Using Jason*. John Wiley and Sons, Ltd, 2007.

[2] A. Casali, L. Godo, and C. Sierra. Graded models for bdi agents. In J. Leite and P. Torroni, editors, *CLIMA V, Lisboa, Portugal*, pages 18—33, 2004.

[3] R. Conte and M. Paolucci. *Reputation in artificial societies: Social beliefs for social order*. Kluwer Academic Publishers, 2002.

[4] A. di Salvatore, I. Pinyol, M. Paolucci, and J. Sabater. Grounding reputation experiments. a replication of a simple market with image exchange. In *Proceedings of the M2M'07, Marseille, France*, pages 32—45, 2007.

[5] F. Giunchiglia and L. Serafini. Multilanguage hierarchical logic (or: How we can do without modal logics). *Journal of AI*, 65:29—70, 1994.

[6] M. Luck, P. McBurney, O. Shehory, and S. Willmott. *Agent Technology: Computing as Interaction (A Roadmap for Agent Based Computing)*. AgentLink, 2005.

[7] S. Parsons, C. Sierra, and N. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, 1998.

[8] I. Pinyol, M. Paolucci, J. Sabater-Mir, and R. Conte. Beyond accuracy. reputation for partner selection with lies and retaliation. In *Proceedings of the MABS'07. Hawaii, USA.*, volume 5003 of *LNCS*, pages 128–140. Springer, 2007.

[9] I. Pinyol, J. Sabater-Mir, and P. Dellunde. Probabilistic dynamic belief logic for image and reputation. In *Proc. of the CCIA'08, Empuries, Spain.*, 2008.

[10] J. Sabater and C. Sierra. Review on computational trust and reputation models. *Artif. Intel. Rev.*, 24(1):33–60, 2005.

[11] J. Sabater-Mir, M. Paolucci, and R. Conte. Repage: Reputation and image among limited autonomous partners. *JASSS*, 9(2), 2006.