Dealing with trust and distrust in Agents Societies

(Extended Abstract)

Elisabetta Erriquez
Department of Computer Science
University of Liverpool
Liverpool L69 3BX, UK

e.erriquez@liverpool.ac.uk

ABSTRACT

Agents in Multi-Agent Systems depend on interactions with others to achieve their goals. Often, goals of agents conflict with each other, and agents can be unreliable or deceitful. Therefore, trust and reputation are key issues in this domain. As in human societies, software agents must interact with other agents in settings where there is the possibility that they can be exploited. This suggests the need for computational models of trust and reputation that can be used by software agents, therefore much research has investigated this issue over the past decade [1, 13, 4, 10, 15, 16, 8, 14].

This thesis concentrates on two important questions, therefore it is divided in two parts. The first question is what sources agents can use to build their trust of others upon. The second question is how agents can use trust and reputation concepts to form stable coalitions.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent Systems; I.2.4 [**Knowledge representation formalisms and methods**]

General Terms

Theory, Design, Experimentation, Performance

Keywords

models of trust, society models, trading competition

1. INTRODUCTION

Autonomous agents use trust and reputation to minimise the uncertainty associated with agent interactions. Usually agents gather and compute trust information from the direct interactions they have with each other. Although direct interactions are the most reliable source of information, information about them may not always be available. Therefore, the agent might not be able to form an opinion, based just on direct experiences, on every agent in the society without running the risk of incurring losses. In the first part we investigate the conjecture that agents who make decisions in scenarios where trust is important can benefit from the use of a *social structure*, representing the social relationships that exist between agents. Section 1.1 presents a description of our approach.

Cite as: Dealing with trust and distrust in Agents Societies (Extended Abstract), E. Erriquez, *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. 1353-1354.

Previous work has utilised the notions of reputation and trust in promoting successful cooperation in Multi-Agent Systems. In open distributed systems, where there are many components that can enter and leave the system as they wish, the notion of *trust* becomes key when it comes to decisions about who to cooperate with and when. In the second part of the thesis we present an abstract framework that allows agents to form coalitions with agents that they believe to be trustworthy. Section 1.2 describes brefly the basis of the framework.

1.1 Social Structure for Trust

The first part of this thesis aim to answer the important question about what sources agents can use to build their trust of others upon. For example, agent a can base his trust or reputation of agent b using experience of previous interactions between the two; or agent a might ask a third party c about its opinion regarding b. An important additional source of trust is to use information about the social relationship (here called the *social structure*) between agents [14]. If a and b are competing for the same resources, for example, this may negatively affect the way they trust each other. Similarly, if agents a and b are likely to have complementary resources, and their cooperation would benefit both, it seems likely that they would be more inclined to trust each other.

Although models of social structure have begun to be considered in models of trust and reputation [14], to date, *implementing* social structures, and hence properly *evaluating* their added value and *validating* them, has not been done. And, most importantly, the issue of how a social structure *evolves* does not appear to have been considered in the literature. These issues are addressed in the first part of this thesis.

In this part, we outline a way to combine concepts of social networking and trust relationships. For the first time, we present empirical evidence that a technique to build and maintain a social network representation of the environment allows a trust model to be more effective in selecting trustworthy agents. Agents use their social structure to obtain knowledge that they could not gather otherwise, and use this knowledge to filter their trust relationships. Although the idea of a social structure had already been presented previously [14], there is no indication of how each agent would build this social network representation. The only attempt made is in [2]. However, the proposed model has never been implemented or validated.

In this thesis, we present a method for agents to build a social network representation of their local environment. Using insight from previous interactions and reputation information, agents can maintain their own representation of such environments. With this extended perception of the environments, agents can make more informed decisions.

We provide an implementation of such concept of social structure and test and analyse the result of the use of such a structure in a trust model. We use the the ART testbed [6] as platform for our tests. The ART testbed was developed in order to compare different models for trust in agent communities, and to provide an experimental standard.

With the approach proposed, we strive towards building an archetypal model for trust by combining the concepts of social networking and trust and reputation relationships.

1.2 An abstract framework for Trust

The second part of this thesis is concentrated on using trust and reputation concepts to help agents to form stable coalitions. In fact, the second important question we concentrate on is how agents can use their trust evaluations on other agents to make decisions about who to form a coalition with.

The goal of coalition formation is typically to form robust, cohesive groups that can cooperate to the mutual benefit of all the coalition members. When Multi-Agent Systems are inhabited by agents with their own objectives, it not only becomes plausible that some agents are not trustable, the consequences of joining a coalition of which some members cannot be trusted, or do not trust each other, becomes a key aspect in the decision of whether or not to join a group of agents.

With a relatively small number of exceptions, existing models of coalition formation do not generally consider trust [3, 9]. In more general models [11, 7], individual agents use information about reputation and trust to rank agents according to their level of trustworthiness. Therefore, if an agent decides to form a coalition, it can select those agents he reckons to be trustworthy. Or, alternatively, if an agent is asked to join a coalition, he can assess his trust in the requesting agent and decide whether or not to run the risk of joining a coalition with him. However, we argue that these models lack a *global* view. They only consider the trust binding the agent starting the coalition and the agents receiving the request to join the coalition.

The second part of this thesis addresses this restriction. We propose an abstract framework through which autonomous, self-interested agents can form coalitions based on information relating to trust. In fact, we use *distrust* as the key social concept in our work. Luckily, in many societies, trust is the norm and distrust the exception, so it seems reasonable to assume that a system is provided with information of agents that distrust each other based on previous experiences, rather than on reports of trust. Moreover, in several circumstances, it makes sense to assume that agents base their decision on which coalition they form on explicit information of distrust, rather than on information about trust. So, we focus on how distrust can be used as a mechanism for modelling and reasoning about the reliability of others, and, more importantly, about how to form coalitions that satisfy some stability criteria. We present several notions of mutually trusting coalitions and define different measures to aggregate the information presented in our model.

Taking distrust as the basic entity in our model allows us to benefit in the sense of deriving our core definitions by analogy with a popular and highly influential approach within *argumentation theory* [12]. Specifically, the distrust-based models that we introduce are inspired by the *abstract argumentation frameworks* proposed by Dung [5]. In Dung's framework, an attack relation between arguments is the basic notion, which inspired us to model a distrust relation between agents. We show that several notions of stability and of extensions in the theory of Dung naturally carry over to a system where distrust, rather than attack, is at the core. We extend and refine some of these notions to our trust setting.

2. REFERENCES

[1] Alfarez Abdul-Rahman and Stephen Hailes. Supporting trust in virtual communities. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, 2000.

[2] Ronald Ashri, Sarvapali D. Ramchurn, Jordi Sabater, Michael Luck, and Nicholas R. Jennings. Trust evaluation through relationship analysis. In *AAMAS '05: Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 1005–1011, New York, NY, USA, 2005. ACM.

[3] Silvia Breban and Julita Vassileva. Long-term coalitions for the electronic marketplace. In *Proceedings of the E-Commerce Applications Workshop, Canadian AI Conference*, 2001.

[4] Cristiano Castelfranchi and Rino Falcone. Principles of trust for mas: cognitive anatomy, social importance, and quantification. In *Principles of trust for MAS: cognitive anatomy, social importance, and quantification*, pages 72–79, 1998.

[5] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *AI*, 77:321–357, 1995.

[6] Karen K. Fullam, Tomas B. Klos, Guillaume Muller, Jordi Sabater-Mir, Zvi Topol, K. Suzanne Barber, Jeffrey Rosenschein, and Laurent Vercouter. The agent reputation and trust (art) testbed architecture. In *Proceeding of the 2005 conference on Artificial Intelligence Research and Development*, pages 389–396, Amsterdam, The Netherlands, The Netherlands, 2005. IOS Press.

[7] Nathan Griffiths and Michael Luck. Coalition formation through motivation and trust, 2003.

[8] Trung Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 13(2):119–154, 2006.

[9] Guo Lei, Wang Xiaolin, and Zeng Guangzhou. Trust-based optimal workplace coalition generation. pages 1 –4, dec. 2009.

[10] Stephen Paul Marsh. Formalising trust as a computational concept. Technical report, 1994.

[11] Zhou Qing-hua, Wang Chong-jun, and Xie Jun-yuan. volume 5, pages 541 –545, aug. 2009.

[12] I. Rahwan and G. R. Simari, editors. *Argumentation in Artificial Intelligence*. 2009.

[13] J. Carbo Rubiera, J. M. Molina Lopez, and J. D. Muro. A fuzzy model of reputation in multi-agent systems. In *Agents*, pages 25–26, 2001.

[14] Jordi Sabater and Carles Sierra. Reputation and social network analysis in multi-agent systems. In *AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 475–482, New York, NY, USA, 2002. ACM.

[15] Michael Schillo, Petra Funk, Im Stadtwald, and Michael Rovatsos. Using trust for detecting deceitful agents in artificial societies, 2000.

[16] W. T. Teacy, Jigar Patel, Nicholas R. Jennings, and Michael Luck. Travos: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12(2):183–198, 2006.