Towards the next generation of computational trust and reputation models

Jordi Sabater-Mir

Artificial Intelligence Research Institute (IIIA-CSIC)
Campus UAB, Bellaterra, Catalonia, Spain
jsabater@iiia.csic.es

The scientific research in the area of computational trust and reputation mechanisms for virtual societies is a recent discipline oriented to increase the reliability and performance of electronic communities by introducing in such communities these well known human social control mechanisms.

Computer science has moved from the paradigm of an isolated machine to the paradigm of a network of systems and of distributed computing. Likewise, artificial intelligence is quickly moving from the paradigm of an isolated and non-situated intelligence to the paradigm of a situated, social and collective intelligence. This new "social" dimension is the main responsible of the increasing interest on computational trust and reputation mechanisms applied to electronic societies.

Computational trust and reputation systems have been recognized as key factors for successful electronic commerce adoption. These systems are used by intelligent software agents both as a mechanism of search for trustworthy exchange partners and as an incentive in decision-making about whether or not to honour contracts. Reputation is also used in electronic markets as a trust-enforcing, deterrent, and incentive mechanism to avoid cheaters and frauds. Another area of application in agent technology is teamwork and cooperation.

However, in spite of the obvious utility of using a trust and reputation system, they are still not a usual element you can find in an agent architecture like it is the case for example of planners or communication modules. Trust and reputation systems still belong to the set of "not so important" elements of an agent architecture. This is something it has to change if we really want artificial socially intelligent entities.

If we analyze which computational trust and reputation models are used nowadays, we will notice that only the simplest models are really used (see for example e-Bay[1]). Till now, the use of complex trust and reputation models has been somehow questioned and reduced to academic environments. Why an artificial agent has to use one of these complex models if it is evolving in a simple community with 10 or 15 agents that have limited interaction capabilities? In other words, the applications and environments are too simple to justify a complex trust and reputation model that only has sense in a socially complex environment.

However this situation is changing very quickly. The increase of the global connectivity (everything is inter-connected anytime and everywhere) is bringing us a new concept of environment where virtual entities need to interact among

them and with humans in a complex way. It is clear then, that in a few years we will start to see the virtual societies we have been theorizing about (and waiting) for a long time in the agents community. In this kind of environment, it will be a must for virtual entities to use complex trust and reputation models if they want to be successful.

Now the question is, are the current trust and reputation models ready to fulfill the requirements of these new virtual societies? We think the answer is no. And this negative answer is because of two reasons: the internals of the model and how the model is being integrated with the rest of the elements of the agent architecture.

Up to now, almost all the efforts have been directed to build trust and reputation models based on a pure numerical approach. There are plenty of these models[2, 3] and the mechanisms they use to calculate the trust and reputation values go from simple aggregation of values[4] to the use of probability theory[5], fuzzy logic[6] or the use of entropy[7] just to put some examples. At the end, each model manipulates the input data in a different way trying to obtain the most accurate trust and reputation values for a given subject. However, if we want to undertake the problems found in socially complex virtual societies, we need theoretically more sophisticated trust and reputation systems. The internal process the trust and reputation system is following to arrive to a final trust and reputation value is as important as the result itself. By ignoring this, we are losing a lot of information that is crucial in order to be successful in the complex world of social relations. We think the solution for that is to use a cognitive approach supported by a solid cognitive theory behind.

The second problem is the integration of the model with the rest of the elements that compound the agent architecture. At this moment, current models are purely reactive black boxes. They receive different inputs (that vary from model to model: direct experiences, witness information, social information...) and, when queried, return the current trust and/or reputation value that has been calculated using the inputs that the model can deal with.

If we really want useful trust and reputation systems, we have to transform them in a proactive element of the agent. The immediate consequence of this proactiveness is that the trust and reputation system could participate in the decision making process by suggesting actions, strategies or complete plans that could help to improve the reliability of the trust and reputation values. But perhaps more important, this new approach opens the possibility of something that till now has been completely ignored, the possibility for the agent to manage its own credibility and reputation in front of the rest of the community. This can only be achieved if the trust and reputation system is at the same level that the other elements that compound the agent architecture and can influence the decision making process by proposing actions and plans to influence the perception others have towards it.

Our current work is going in these two directions: improve the internals of the trust and reputation systems by using a cognitive approach and at the same time analyze how the trust and reputation system can be integrated in the

agent architecture to provide the functionality required by a socially intelligent virtual entity. We think these are the two keystones for the next generation of computational trust and reputation models. Our first step on this direction is RepAge[8], a computational reputation model based on a cognitive theory of reputation[9] and a previous trust and reputation model called ReGreT[10].

References

1. eBay: eBay, http://www.eBay.com. (2002)
2. Sabater, J., Sierra, C.: Review on computational trust and reputation models. Artificial Intelligence Review **24** (2005) 33–60
3. Ramchurn, S., Hunyh, D., Jennings, N.: Trust in multi-agent systems. The Knowledge Engineering Review **1** (2004) 1—25
4. Zacharia, G.: Collaborative reputation mechanisms for online communities. Master's thesis, Massachusetts Institute of Technology (1999)
5. Schillo, M., Funk, P., Rovatsos, M.: Using trust for detecting deceitful agents in artificial societites. Applied Artificial Intelligence (2000) 825—848
6. Carbo, J., Molina, J., Davila, J.: Trust management through fuzzy reputation. Int. Journal in Cooperative Information Systems **12** (2002) 135—155
7. Sierra, C., Debenham, J.: An information-based model for trust. In: Proceedings of the fourth international joint conference on autonomous agents and multiagent systems (AAMAS-05), Utrecht, Netherlands. (2005) 497—504
8. Paolucci, M., Sabater-Mir, J., Conte, R.: What if? dealing with uncertainty in repage's mental landscape. In: Proceedings International Conference on Intelligent Agent Technology (IAT-05), Compigne, France. (2005) 372—378
9. Conte, R., Paolucci, M.: Reputation in artificial societies: Social beliefs for social order. Kluwer Academic Publishers (2002)
10. Sabater, J.: Trust and reputation for agent societies. PhD thesis, Universitat Autonoma de Barcelona (UAB) (2003)