USING TRUST FOR DELEGATION IN MULTI-AGENT SYSTEMS

Ioan Alfred Letia
Radu Razvan Slavescu

TechnicalUniversty of Cluj-Napoca
Department of Computer Science
Baritiu 28, RO-400391, Cluj-Napoca, Romania
{letia, srazvan}@cs-gw.utcluj.ro

Abstract: In this paper, we present a new approach for numerical trust evaluation, based on it's cognitive components of competence and willingness and how the willingness could be seen as a measure for the price of cooperation . We also show how this approach could be extended to cover the case of re-delegation, i.e. delegating a task from one agent to an intermediate one and then to the agent who actually does the job. We study how trust will be distributed among the nodes of a social network, especially in the situation when an agent possesses only partial information about the delegation chain topology and about the competence and willingness of the intermediate agents.

Keywords: artificial intelligence, agents, cooperation

1. INTRODUCTION

Cooperation proved to be of paramount importance in the area of multi-agent systems. In the vast majority of situations, one agent alone is not able to accomplish a complex task, due to its limited abilities and/or resources. Therefore, the agent will have to rely on some other agents in order to have parts of the aimed goal achieved. When delegating tasks to the others, the main problem which arises is choosing partners such that the newly created team is actually able to fulfill the goal. Hence, the importance of modeling and understanding the delegation mechanism in detail, as well as the capabilities of potential partners.

When employing someone else to act on behalf of himself, an agent has to solve some challenges. The most serious one is that such an action requires a certain price to be paid, i.e. agents in real systems are not committed and perhaps not even able to cooperate unconditionally. Committing to pay for cooperation involves a certain amount of risk for the agent who makes he pay. Thus, an agent must obtain the optimal trade-off between the risk involved and the price he or she has to pay.

For this problem, one solution which proved robust and scalable is based on the concept of trust. This is seen as an attitude an agent has towards the future evolution of events. The trust an agent (the trustor) has in another agent (the trustee) will be the result of the trustee's previous behavior. In this approach, trust describes the history of the cooperation relationships between the trustor and the trustee relative to a specific type of goal.

The approach suggested by this paper tries to emphasize the concept of trust in its social dimension, taking into consideration multiple delegations among agents. By this, we mean the following: if an agent is delegated a task, he could decide to delegate it further. Therefore, a "delegation chain" could be established: a sequence of agents along which a specific task is delegated till accomplished. In this case, trust will describe a cooperation chain rather than a cooperation relationship. Since not all interactions along this chain are known by the initial trustor, trust evaluation should be based strictly on the information available to him. We investigated how this phenomenon influences the trust distribution at the level of the agent society.

Our work focused on two goals:

- developing a new approach for a numerical evaluation of trust; this aims to model in more detail the perceived risk of such a cooperation and to estimate to what extent such a risk could be dealt with;
- investigating inter-agent cooperation through delegation chains, when the delegated task is passed further several times; in such a situation, trust assessment must rely only on partial information on the delegation chain topology and on the level of competence and commitment of the other agents involved.

The paper is organized in the following manner. Section 2 positions our work in the context of similar achievements in the field of trust modeling and its applications to delegation. In Section 3, we describe in detail our model for assessing trust, based on a cognitive approach of the notion. Section 4 investigates the trust-based delegation and points out the advances it makes compared to the existing approaches. Section 5 presents the test scenario we used for experiments and discusses the results we have obtained so far. Section 6 concludes and sketches the possibilities of development.

2. RELATED WORK

The approach presented in (Gambetta, 1991) regards trust as a subjective probability with which an agent assesses the perspective of another agent or group of agents to perform a particular action, both before he can monitor such an action (or independently of his capacity ever to be able to monitor it) and in the context in which it affects its own actions.

Another approach of Mui (Mui, 2002; Mui *et al.* 2002) describes the same notion as a subjective expectation an agent has about another's future behavior based on the history of their encounters. This definition insists on the fact the evidence on which an agent might rely when evaluating this subjective perspective (i.e. trust) is provided by the sequence of prior experiences he had with the agent he aims to assess. Jonker and Treur (1999) suggest how such an assessment could be accomplished and states the properties an evaluation function should have.

Paper (Grandison and Sloman, 2000) shows trust is a context dependent, subjective, asymmetric binary relation, with a value and a temporal dimension.

Marsh (1994) presents one of the earliest approaches for evaluating trust; his work reveals some key components of trust, such as the importance of situation, the perceived risk and the perceived competence.

Work presented in (Zacharia, 1999) describes Sporas a reputation measurement functions for agent societies. In (Sabater and Sierra, 2002) REGRET, a system for measuring reputations in gregarious societies, is introduced and its performance is compared to the one of Sporas.

In (Castelfranchi and Falcone, 1998; Falcone and Castelfranchi, 2004), a cognitive model of trust is presented. According to this model, an agent's decision of delegating a task to another agent is the result of the trust the first agent has in the second one. Trust is seen as "the mental counter-part" of delegation (Castelfranchi and Falcone, 1999). The same papers describe the beliefs involved in trust. Paper (Falcone *et al.*, 2004) suggests an approach for simulating them in the context of task delegating among agents.

Following the line of the cognitive model of trust, (Falcone and Castelfranchi, 1998) elaborates a theory of the delegating process in case of multi-agent systems. A taxonomy of delegation is developed, which takes into consideration the way (active/passive) the goal is achieved and the presence/absence of mutual beliefs. Delegation agreement is when there exists mutual belief and mutual active achievement of goal

A model for delegation which makes use of modal logics is developed in (Norman and Reed, 2004). Delegation is formalized in the context of imperative communication acts. Different types of delegation are described, together with their corresponding responsibility holders.

3. ASSESSING TRUST

Our approach relies on the idea that trust and delegation are tightly connected. This section presents our approach for assessing trust, which is one of the contributions made by this paper. In the next section, we describe how a simulation of delegation and of task accomplishment is conducted, as well as how an agent's beliefs are modified over time.

Papers (Castelfranchi and Falcone, 1998) and (Falcone and Castelfranchi 2004) conceptually describe a cognitive model of trust. Elementary believes involved in it are enumerated. Firstly, the agent trusting another one must be a cognitive agent, i.e. an agent endowed with *goals* and *beliefs* because:

- an agent trusts another only relatively to a goal *g*, which might temporarily become the trustee's goal;
- trust itself consists of *beliefs*, being an attitude of the trusting agent towards the

agent he trusts, concerning action α, which is relevant for the goal g.

If we consider an agent x, having a goal g that wants to delegate to agent y, the following beliefs are enumerated for agent x:

- Competence: x believes that y is able to perform a certain task, so that it could play a role in x's plan;
- Willingness: x believes not only that y is able to perform the task, but he also believes that y will actually do what x wants;
- Dependence: x believes either that x depends on the performance of g by y or that it is better for him to rely than not to rely on y;
- Fulfillment: x believes that g will be achieved (without necessarily carrying about the specific way in which this will happen);
- Willingness: x believes that y has decided and intends to do α. Therefore, trust needs modeling the mind of the other;
- Persistence: x believes also that y is stable enough in his intentions, i.e. y is not likely to change his commitment about performing α.
- Self-confidence: x believes that y knows that y can do α.

The competence and willingness beliefs are seen as “the cognitive kernel of trust". The model employs in our work uses a simplified approach, focusing solely on these two beliefs. Competence is a measure of agent's physical capabilities to carry a task. It could be considered constant and independent of the delegating contract terms and parts. Willingness estimates an agent's commitment to have the task accomplished, e.g. the effort or number of tries she wants to spend in order to fulfill the goal. Its level could be subject of negotiations between the trustor and the trustee (or, equivalent, between the agent who delegates the task and the agent to whom the task is delegated). If the trustor wants to have it increased, perhaps he will have to pay a price for this to the trustee and this will become a part of their delegation contract. Starting from the approach in (Falcone *et al.*, 2004), we elaborated a way to simulate competence and willingness. We assumed the interaction is *always* based on a prior agreement whose terms can be controlled by the trustor. We will see later the agreement actually states the willingness level an agent aims to ensure before delegating a task to a partner.

In our approach, competence is a real number in [0,1] which describes the agent ability to do some task. It is constant for a specific agent and a specific task, but not necessarily known by the partners. The higher the competence an agent is endowed with, the higher the number measuring it in the model.

Willingness is modeled by a strictly positive natural number. Willingness depends on the contract between the trustor and the trustee. A higher willingness could determine, for instance, a higher price the trustor has to pay in order to secure the task accomplishment by the trustee. The semantics of the willingness is different for intermediate and for task accomplishing agents. For an agent who is the final link in a delegation chain (i.e. an agent who actually does the job) the willingness means the number of tries he makes in order to have the task accomplished. For an intermediate agent, it means either the number of different agents to whom he passes the task or, if this is not possible, the number of times it delegates the task to another agent.

When a task needs to be carried, the following simulation takes place: a random number in [0, 1] is generated, according to a normal probability distribution. The values for μ and σ are fixed for a given agent and task and are tunable parameters of the experiment. After such a number is generated, the corresponding agent decides to report the value of the experiment. The reported value is not necessarily equal to the generated one, but might be less (taken from a distribution with a different μ, corresponding to a failure of the delegated task). The decision about which value is reported belongs to the delegated agent and depends on the delegator (perhaps the delegatee wants to help certain agents more than others).

We describe now how this is modeled from an agent's point of view. First, we assume such an agent pursues a certain goal, which he wants to be achieved by delegation. Then, we assume the agent is a cognitive one, i.e. it is endowed with beliefs of competence and willingness. When modeling this, we have to specify how a value is attached by the trustee to each of the two beliefs. For each known agent, the trustee will compute a trust value. This value is always in (0, 1) and has an initial value specified as a parameter of experiment (e.g. 0.5, corresponding to an average trust level). It is adjusted after each success / failure of a contract, according to equation (1):

$$T' = \gamma * T + (1 - \gamma) * (A-P) \quad (1)$$

where:
T' is the new trust level
T is the current trust level
A is 1 if the current experience is greater than τ and 0 otherwise
τ is a threshold value the trustor wants to secure
P is the expected value of the current contract, computed as explained below
γ is a weight factor which measures the impact of the current experience over the trust value

For computational convenience, T is always adjusted to a minimum value of 0.001 if the value computed according to equation (1) falls below this. The same holds for the case when T exceeds 1; in this situation, T will get the maximum value of 0.999.

The trust level is used by the trustor to estimate the competence level he should expect from the trustee. This way, in round *n* of testing, competence is estimated as the average of all prior trust levels till round *n*.

When making a contract, the trustor will compute its (subjective) success likelihood, knowing the simulation model involving competence and willingness described above. This actually represents a trust regarding future (as opposite to the value *T*, which represents the history-based trust):

$$P = 1 - (1 - C)^{W} \quad (2)$$

Where:
P is the likelihood of success for the contract
C is the perceived competence (estimated by an average of the last reported results)
W is the willingness level established by the contract

For each kind of situation, the trustor must have a risk threshold τ. This represents the minimum value the trust representing future must have for the agent to rely on the trustee. It could be dictated, for example, by the importance the trustee estimates for the current situation. In this case, in order to make the delegating decision, the trustor must make sure the negotiated willingness is at least equal to the one described by equation (3):

$$W = \frac{\ln(1-\tau)}{\ln(1-C)} \quad (3)$$

where
W means the minimum willingness level the contract should secure in order to rely on the hypothesis of success
τ is the success threshold
C is the perceived competence

When delegating a task to an agent, the willingness value could be regarded as the amount the trustor has to pay in order to make sure the task is accomplished, given his prior knowledge about the results of similar accomplishments in the past. This requires the value of willingness to be computed according to equation (3).

In order to study the performance of this approach, we conducted an experiment similar to that presented in (Zacharia, 1999) and reproduced in (Sabater and Sierra, 2002). An agent behaves reliably in the first 1/3 of the experiences, with an output level of performances of 2400 (on a scale form 0 to 3000, as in Sporas). Then, for the next 2/3, its performance drops to 900. The ratings received by the agent are taken from a normal probability distribution having μ equal to its actual performance and σ equal to 300. We investigate the algorithm's capability of adjusting the computed trust value to the actual performances of the agents.

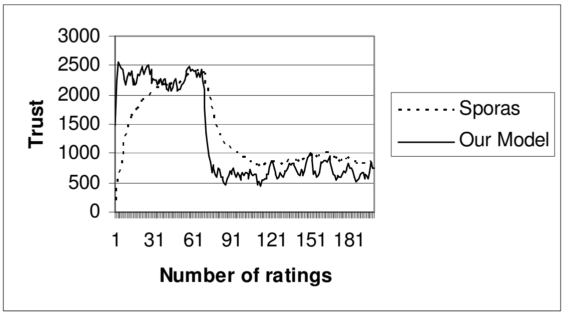

Fig. 1. Abuse of prior performance

The results obtained for γ=0.95, τ=0.75, initial value of competence 1500 are summarized in Figure 1. As one can see, the cognitive approach presented reacts to the changes in an agent's behavior faster compared to Sporas (where we had θ=8, σ=200). Similar results have been obtained when comparing the algorithm with REGRET reputation mechanism.

4. COGNITIVE TRUST AND DELEGATION

Work in the field of theory of delegation tries to develop a formal framework for the actions and options an agent has in the open world he lives. In paper (Norman and Reed, 2004), the following axioms are stated about delegation:

$$S_x\, S_y\, A \rightarrow S_x\, A \quad (4)$$
$$S_x\, T_y\, \alpha \rightarrow T_x\, \alpha \quad (5)$$

They basically explain that if agent *x* sees to it that agent *y* sees to it that action α is done, then *x* could be held responsible for α to be done. This property, combined with the inter-definition between obligation and permission:

$$Pp \leftrightarrow \neg\, O \neg\, p \quad (6)$$

leads to the conclusion that, in multi-agent systems, multiple delegation is possible if not expressly forbidden. However, in order to avoid the problem of circular delegation, we limited the depth of delegation chain to 2 (i.e. a delegated task could be re-delegated only once).

Even if we assumed the interaction is based on a prior agreement whose terms are on the trustor's control, we must distinguish carefully between such an agreement and the actual task performance. From the trustor point of view, having a task accomplished means making a single agreement. In fact, this could mean more than one agreement, since it is possible that the trustee delegates the task further to another agent. In this case, the trustee makes a separate agreement with the latter agent, i.e. makes a trust-based choice at his turn, building a delegation chain. However, we must emphasize the process is fully transparent for the first agent. We either have the trustee *E* actually performing the task delegated by *A* or we might have agent *I* as an intermediate agent which may pass the task received from agent *A* to another agent *E'*. Decision whether agents *E* and *E'* are different or identical is completely transparent for agent *A* and belongs strictly to agent *I*. *A* does not even know how many agent are in the chain behind *I*; the only agent responsible with the task, in his opinion, is agent *I*.

This paper claims the following assumption should be made about the trust adjustment after a contract end: each agent will adjust its own trust level concerning the trustee he was involved in contract with, regardless the actual structure of the delegation chain.

The advantage of this approach is it offers a simple way to cope with the complex cooperation relationships which might appear among agents. Unlike the trust metrics which allow an agent to select trusted *partners* for cooperation, our approach allow agents to select trusted *relationships*, which models not only physical abilities, but also social interaction abilities. By employing trust for partner selection, we get a way for selecting the best relationship among the plethora of the cooperation relationships which might exist in a complex multi-agent system.

5. EXPERIMENT SCENARIO

A multi-agent system has been implemented, using a set of C++ classes and the LAM tool. A limited number of simulations have been done; extensive experiments are to be conducted.

Agents involved in our test assess their partners based on direct experiences: each agent picks a partner and delegates him a specific task. One agent can refuse to accomplish a task if it is not able to do it or if he is not allowed to delegate it further due to re-delegation forbearance. After completing the task, the delegating agent evaluates the result and updates his trust level into the partner accordingly. Each time a cooperation partner is needed, one of the evaluated agents is appointed to this task. As previously stated, re-delegation is allowed only once. After each experience, the trust level is modified according to equation (1). No agent ever gets banned even if steadily distrustful.

Once an experience completed, it gets evaluated, with a value in {0, 1}, corresponding to failure and success respectively. The adjustment takes into account the agent's recorded competence, stated willingness and predicted success probability.

The experiments aimed to study the possibility of multiple delegation and the cooperation relationships it might generate. For instance, if *A* needs a task to be accomplished and it delegates it to *I*, the latter might delegate the task further to agent *E*.

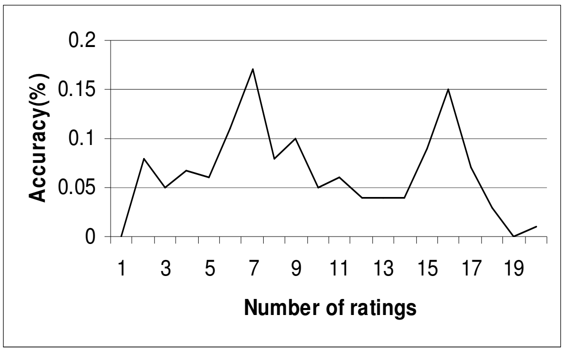

Fig. 2. Delegation accuracy

The credit for task accomplishment, as well as the blame for a failure is given entirely to *I*. Therefore, trust values *A* possesses will be influenced both by the capacity of the others to accomplish it and by the capacity of them to find another reliable partners. It will describe cooperation networks rather than isolated agents, which we consider to be more realistic from the perspective of the open world of agents.

One of the ideas this paper aims to stress is: one important function of trust is delegating in order to have one task accomplished. But this could mean multiple delegation and building a "delegation chain" which is a sequence of agents along which a specific task is delegated till accomplished. This offers the opportunity of consolidating a cooperation path in the agent's social network. Visualizing and studying them was one of the goals of our work. Let us consider the following example: agent *A* needs to have task *T* done, but cannot do it by himself. One option is to delegate it to agent *E* and let *him* accomplish it. The other option is to delegate the task to agent *I* and expect him to do the job; but *I* cannot do it either and delegate it further to the same agent *E*. The question is: in the latter case will the chances of success be higher? We believe the answer is “yes" as it is possible for *I* to have better knowledge about *E* than *A*. Since no general monitoring system is likely to be available, an agent should take advantage of the other's local knowledge about the partners, while making sure the accuracy of delegation is still offered by direct partner assessment.

We conducted a set of preliminary experiments in order to test the model's validity. We compared the accuracy of delegation decision for the case when re-delegation is forbidden, respectively allowed. Figure 2 depicts the difference between the accuracies of delegating decisions in the two cases. The results show a better accuracy for the re-delegation case. We believe this is especially the case when a lower number of mutual interactions are available; in this particular type of situations, the knowledge about the social connections of the other get a higher importance.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we sketched a method for assessing trust based on a cognitive model of this notion. We also presented its possible application to the problem of delegation in multi-agent systems. The trust assessment method relies on the cognitive notions of goal and believes, especially on those concerning competence and willingness; it also offers the opportunity to simulate them in quite an intuitive manner. The model presents some properties which appear natural to us, especially that of applying a reward/penalty related to the expected gain or loss. More than that, it allows an agent to compute how much she should invest in order to secure a certain threshold of success of its goal.

In the same time, the model aims to regard trust more as a societal concept, describing cooperation networks, rather than binary cooperation relationships. It copes with delegation chains; this enables the trustor to exploit the existing social knowledge about its peers and the cooperation relationships which are already present in the system, aside from knowledge about their capabilities alone.

The model never bans cooperation; this makes it possible to avoid possible deadlocks and the issues related to the intransitivity of distrust. Deciding with whom to cooperate is up to the agent; deciding to cooperate if necessary is stated by the system and cannot be overturned by a specific agent.

One direction we want to investigate is to combine this model with a recommendation system. When an agent decides it should delegate a certain task but has no knowledge about potential partners, she might use a recommendation system. This means it should use the knowledge of the other agents in order to discover a potential partner. Our model should operate a distinction between responsibilities involved by a recommendation versus a those involved by a multiple delegating chain. One agent might be skillful at making recommendations, but poor at delegating or solving tasks by himself.

ACKNOWLEDGEMENTS

The work for this article has been supported in part by the National University Research Council in Romania, within the framework of the projects number 98/2003 and 240/2004.

REFERENCES

Castelfranchi, C. and R. Falcone (1998). Principles of trust for MAS: Cognitive anatomy, social importance, and quantification. In: *Proceedings of the 3rd International Conference on Multi-Agent Systems (ICMAS-98), Paris, France*.

Castelfranchi, C. and R. Falcone (1999). The dynamics of trust: from beliefs to actions.

Falcone, R. and C. Castelfranchi (1998). Towards a theory of delegation for agent-based systems. In: *Robotics and Autonomous Systems, Special issue on Multi-Agent Rationality*. Elsevier.

Falcone, R. and C. Castelfranchi (2004). Trust dynamics: How trust is influenced by direct experiences and by trust itself. In Proceedings of AAMAS04.

Falcone, R., G. Pezzulo, C. Castelfranchi and G. Calvi (2004). Why a cognitive trustier perform better: Simulating trust-based contract nets. In: *Proceedings of AA-MAS04*.

Gambetta, D. (1990). *Trust*. Basil Blackwell.

Grandison, T. and M. Sloman (2000). A survey of trust in internet applications. *IEEE Communications Surveys*, **vol. 3(4)**.

Jonker, C. M. and J. Treur (1999). Formal analysis of models for the dynamics of trust based on experiences. In: *Proceedings of the International Workshop on Deception, Fraud and Trust in Agent Societies*, Seattle, WA, USA.

Marsh, S (1994). *Formalizing Trust as a Computational Concept. PhD thesis*, Department of Mathematics and Computer Science, University of Stirling, UK.

Mui, L (2002). *Computational models of trust and reputation: Agents, evolutionary games, and social networks . MIT PhD thesis*.

Mui, L., M. Mohtashemi and A. Halberstadt (2002). A computational model for trust and reputation. In: *Proceedings of the 35th Hawaii International Conference on System Sciences*, *Hawaii, USA*.

Norman, T. J. and C. Reed (2002). A model of delegation for multi-agent systems. *Foundations and applications of Multi-Agent Systems*, **LNAI 2403**. Springer-Verlag.

Sabater, J. and Sierra C. (2002). Regret: a reputation model for gregarious societies. *First International Joint Conference on Autonomous Agents and Multi-Agent Systems*.

Zacharia, G. (1999). Collaborative reputation mechanisms for online communities. *MIT Master's Thesis.*