Enhancing Trust-Based Competitive Multi Agent Systems by Certified Reputation (Short Paper)

Francesco Buccafurri, Antonello Comi, Gianluca Lax, and Domenico Rosaci

University of Reggio Calabria, Via Graziella, loc. Feo di Vito, 89122 Reggio Cal., Italy
{bucca,antonello.comi,lax,domenico.rosaci}@unirc.it

Abstract. In the past, the experience of the ART community highlighted that, in absence of information about the quality of the recommendation providers, it is better to exploit only the direct knowledge about the environment (i.e., a reliability measure), missing the reputation measure. However, when the size of the agent space becomes large enough and the number of "expert" agents to contact is small, the use of just the reliability is little effective. Unfortunately, the largeness of the agent space makes the problem of the trustworthiness of recommendations very critical, so that the combination of reliability and reputation is not a trivial task. In this paper, we deal with the above problem by studying how the introduction of the notion of certified reputation, and its exploitation to combine reputation and reliability, can improve the performance of an agent in a competitive MAS context. We analyze different populations, using the standard platform ART, highlighting a significant positive impact and providing very interesting results.

1 Introduction

In the context of competitive multi-agent systems (MAS), the term *reliability* indicates a subjective measure of trust, while the term *reputation* is used for a measure of the trust the whole community perceives with respect to a given trust. The reputation assumes a very important role when an agent a does not have a sufficient knowledge about another agent b. Then, a can exploit b's reputation to decide whether b is a reliable interlocutor or not. Several models for competitive MASs have been proposed in the past [7, 10] for representing both reliability and reputation, and in 2004 the platform *Agent Reputation and Trust* (ART) [1] has been created, in order to provide a common framework for homogeneously comparing different trust and reputation models. Using this framework, three ART competitions were organized in 2006, 2007 and 2008, respectively. Several feedbacks have been provided about the limitations of ART, and some of them have been described in [3]. The main evidence is related to the marginal role of the reputation in the tested conditions. For example, [11], describing the agent IAM winner of the first two ART competitions, clearly stated that the IAM agent does not require reputation values from other participants, but it

R. Meersman et al. (Eds.): OTM 2012, Part II, LNCS 7566, pp. 855–862, 2012.

relies only on the variance estimator to estimate the performance of an opinion provider. Also the agent UNO2008 [8], winner of the 2008 ART competition decided not to use the reputation, basing its strategy only on the knowledge that it directly learns from other agents. The approach used by UNO2008 outperformed IAM, but also it overcame other agents that used reputation information besides reliability, like, for example the agent RRAF proposed in [2, 9]. However, the experimental evaluation run on ART considers only scenarios where the number of agents is limited and, consequently, the percentage of expert agents is forced to be appreciable. But, what happens when the number of agents grows in such a way that finding an expert agent becomes really difficult? We argue that the sole reliability does not give satisfactory results, due to the low probability of contacting expert agents. We could thus resume the reputation measure, in addition to reliability. But, the largeness of the agent space makes the problem of the trustworthiness of recommendations highly critical, so that the combination of reliability and reputation in this new scenario becomes a crucial problem.

In the literature, this problem has been theoretically faced [4–6], but, to the best of our knowledge, no effective solution implemented in the context of competitive MAS still exists. In particular, the approach presented in [4] introduces a model of trust allowing agents to actively provide third-party references about their previous performance as a means of building up the trust in them of their potential collaborators. Instead, in [6] some research challenges for trust and reputation systems are discussed, and a research agenda for developing sound and reliable robustness principles and mechanisms for trust and reputation systems are presented. However, these approaches only deal with possible solutions to make "secure" the recommendations provided by the agents, without studying how a "certificated" reputation impacts in competitive multi-agent systems, where it is necessary to use reputation measures combined with reliability measures, taking into account that the agents are in competition among them.

The purpose of this paper is studying whether the introduction of a notion of *security level* of recommendations and its suitable exploitation for the combination of reputation with reliability, can be fruitful in the context of competitive MAS. We remark that the core of the proposal is not the definition of a security protocol, but the analysis of its influence in the competitive MAS setting. To do this, we have performed two campaigns of experiments that show clearly the advantages introduced by our approach. Our agent (called CAST – *Competitive Agent Security in Trust*), compared with the winners of the three past ART competitions, and also with RRAF [2, 9], obtains always the best scores. In particular, we have observed the best performances of CAST in presence of a high-sized population characterized by a low number of *expert* agents. It is worth remarking that the contribution of our work relies its significance on the practical relevance of the assumptions we do about the scenario. Indeed, more and more the application contexts where competitive MAS can be applied (e.g., electronic commerce) regard open and hostile environments with large populations. From this perspective, our results fully preserve the validity of the existing

approaches, but give a significant improvement of the state-of-the-art under the above conditions.

The remaining of the paper is organized as follows. We present the CAST trust model in Section 2. Then, in Section 3, we describe the experiments we have performed to compare our proposal with the state of the art, highlighting advantages and limitations. Finally, in Section 4, we draw our conclusions.

2 The CAST Model

Our model deals with a multi-agent system, whose agents can provide a set of services belonging to a *set of categories SC*. We suppose that when a client needs a service falling into a given category *cat*, it sends a request to the Agent System Manager ASM of the multi-agent system which assigns each request to an agent at pre-determined temporal steps, based on the effectiveness shown by all the agents in the past. When the assignment is done, the client must pay a given price *sp* to the selected agent in order to obtain the service. During each temporal step, the agents of the community can interact with each others, in order to exchange information. We denote by AS the list containing all the agents belonging to the multi-agent systems, and by a_i the i-th agent of AS. The interactions between agents (both collaboration requests and recommendation requests) are executed following a *security protocol* which (possibly) allows the interlocutors to mutually exchange a *proof* synthetically describing the interaction.

In this paper, we do not deal in depth with the issue of the security protocol, since this aspect is fully orthogonal to the core of the proposal in the sense that, in principle, we could choose any security protocol able to produce, as a final state, a [0..1]-real *security level* l (where 1 is the maximum level of trustworthiness), representing a measure of the trustworthiness of the recommendation that a requested agent c provides to a requester agent a about a third agent b. Anyway, we give a very short sketch about how this security level can be obtained in our protocol, in order to make plausible the overall proposal.

We can imagine that the security level is obtained by evaluating the proof that c can show to a in order to guarantee the reliability degree of the provided recommendation. The maximum reliability degree (and, thus, the highest security level) corresponds to the case in which all the transactions occurred between b and c, on which c produces its recommendation, are traced through messages whose authenticity and non-repudiation are based on a Third-Trusted-Part-granted certification. In contrast, the presence of some transaction traced through a weaker mechanism (for example, in our model the transactions can be traced just on the basis of some randomly chosen witnesses), reduces the security level of the recommendation, until the minimum value corresponding to the case of all transactions with no proof.

A set of four mappings, denoted by SR_i, R_i, β_i, and P_i is associated with each agent a_i, where each mapping receives an agent j and a category *cat* as input and yields as output a different trust measure that the agent a_i assigns to the agent a_j, in relation to the category *cat*. Each trust measure is represented by

a real number belonging to the interval $[0, 1]$, where 0 (1, resp.) is the minimum (maximum, resp.) value of trust. In particular: (i) $SR_i(j, cat)$ represents the *service reliability* that the agent a_i assigns to the services provided by the agent a_j for the category cat; (ii) $R_i(j, cat)$ represents the *reputation* that the agent a_i assigns to the agent a_j for the category cat; (iii) $\beta_i(j, cat)$ represents the *preference* that the agent a_i assigns to the usage of the reliability with respect to the reputation in evaluating the agent a_j for the category cat;(iv) $P_i(j, cat)$ represents the overall preference that the agent a_i assigns to the agent a_j for the category cat, based on both the reliability and reputation perceived by a_i.

We also define a *recommendation* as a pair $r = \langle v, l \rangle$, where v and l are two [0..1]-real numbers called *recommendation value* and *recommendation security level*, respectively. All the recommendations received by a_i are stored into a mapping $RECC_i$, that receives two agents j and k and a category cat as input, and yields as output the recommendation $RECC_i(j, k, cat)$ representing the recommendation that the agent a_j provided to the agent a_i about the agent a_k for the category cat. Recall that, according to the definition of recommendation, $RECC_i(j, k, cat)$ is a pair whose elements are denoted by $RECC_i(j, k, c).v$ (recommendation value) and $RECC_i(j, k, c).l$ (recommendation security level), respectively.

The mappings above are updated by the agent a_i at each step, as follows:

Phase 1: Reception of the Recommendations. The agent a_i receives, at the current step, some recommendations from the other agents, in response to previous recommendation requests. These recommendations are stored in the $RECC$ mapping.

Phase 2: Computation of SR mapping: ASM sends a_i the feedbacks for each service s provided in the past step, where the contributions given by other agents to a_i are evaluated. These feedbacks are contained in a mapping $FEED$, where each feedback $FEED(s, j)$ is a real number belonging to $[0, 1]$, representing the quality of the collaboration that the agent a_j provided to the agent a_i concerning the service s. Basing on these feedbacks, the agent a_i updates the mappings SR. Specifically, we choose to compute the current reliability $sr(j, cat)$ shown by an agent a_j in its collaboration with a_i by averaging all the feedbacks concerning a_j. At each new step, this current reliability is taken into account for updating the element SR_i, by averaging the value of SR_i at the previous step and the current reliability computed at the new step. Thus: $SR_i(j, cat) = \alpha \cdot SR_i(j, cat) + (1 - \alpha) \cdot sr(j, cat)$, where α is a real value belonging to $[0, 1]$, representing the importance that a_i gives to the past evaluations of the reliability.

Phase 3: Computation of R and β. The recommendations contained in the mapping $RECC_i$ are used by the agent a_i to compute the reputations of the other agents of the community. In particular, a_i computes the reputation of another agent a_j as a weighted mean of all the recommendations received from the other agents of the community concerning a_j, where the weight of each recommendation value is the corresponding security level. Thus: $R_i(j, cat) = \frac{\sum_{k \in AS, k \neq i} RECC_i(k,j,cat).v \cdot RECC_i(k,j,cat).l}{\sum_{k \in AS, k \neq i} RECC_i(k,j,cat).l}$. The computation of the

average security level of the recommendations related to an agent a_j in the category cat, denoted by $\beta_i(j, cat)$, is obtained by simply averaging the security levels associated to all the recommendations related to a_j in the category cat. Thus: $\beta_i(j, cat) = \frac{\sum_{k \in AS, k \neq i} RECC_i(k,j,cat).l}{|AS|-1}$.

Phase 4: Computation of P. The agent a_i finally computes the overall preference measure $P_i(j, cat)$ in the agent a_j by taking into account both the service reliability $SR_i(j, cat)$ and the reputation $R_i(j, cat)$, using the value of the mapping $\beta_i(j, cat)$ to weight the importance of the service reliability. Formally: $P_i(j, cat) = \beta_i(j, cat) \cdot SR_i(j, cat) + (1 - \beta_i(j, cat)) \cdot R_i(j, cat)$.

At each step, the agent a_i exploits the mapping P to select the most suitable candidates to require a collaboration, paying a price cp for each received contribution.

3 Evaluation

In this section, we describe the experimental campaign aimed to evaluate the advantages and the limitations introduced by CAST. All the material needed for reproducing the experiments is available at `http://www.unirc.it/firma/cast.html`. In our experiment, we used the **A**gent **R**eputation and **T**rust (ART) platform [1], which has a significant acceptation in providing a tool for computing fair comparisons of trust models. On ART, each agent takes the role of an art appraiser who gives appraisals on paintings presented by its clients. In order to fulfill its appraisals, each agent can ask opinions to other agents. These agents are also in competition among them and thus, they may lie in order to fool opponents. The game is supervised by a *simulator* running in a synchronous and step by step manner.

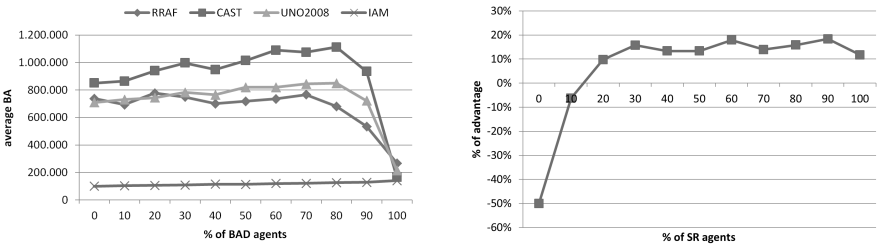
The clients require opinions on paintings to the appraiser agents. The application domain is the "paintings appraisal". Each painting belongs to an *era*, therefore the set of the categories SC is the set of all possible eras. Each agent has a specific expertise level in each era, assigned by the simulator. The error made by an agent while appraising a painting depends on this expertise and the price the appraiser decides to spend for that appraisal. An agent's expertise, defined as its ability to generate an opinion about the value of a painting, is described by a normal distribution of the error between the agent's opinion and the true painting value. The true values of paintings presented by clients are chosen from a uniform distribution known only to the simulation. Likewise, the eras which paintings belong to are also uniformly distributed among the set of eras. Each agent can obtain recommendations about another agent by other players. Each recommendation has a given price rp, which in our experiments has been set to the value 0.01. The agent with the least average relative appraisal error achieves the highest preliminary client share. Finally, the actual client share of each agent a_i takes into account the appraisers client share from the previous step.

In the experiments, we have chosen for the parameters sp and cp the same values used in the ART competition 2008, i.e. $sp = 100$ and $cp = 10$. The

purpose of our experiments is to compare the performances of agents exploiting the trust model here presented, and three other approaches, namely UNO2008 [8], which is the winner of the ART 2008 competition and uses only the reliability in its trust model, RRAF [2, 9], which is an approach exploiting both reliability and reputation, weighting the importance of the reliability with respect to the reputation by a fixed coefficient β, and IAM [11], which is the winner of the ART competitions in 2006 and 2007.

In order to simulate different types of agent populations, we have built, besides the agent CAST, UNO2008, RRAF and IAM that are the subjects of our comparison, three types of dummy agents:(i) SR (Secure Random) agent, provided with a random expertise $e \in [0..1]$, generated by a uniform distribution, responding to a recommendation request with a certified recommendation, whose security level $l \in [0..1]$ is generated by a uniform random distribution; (ii) BAD agent, provided with a random expertise $e \in [0..0.2]$, generated by a uniform distribution, responding to a recommendation request with a certified recommendation having security level generated by a uniform random distribution; (iii) USR (Unsecure Random) agent, provided with a random expertise $e \in [0..1]$, generated by a uniform distribution, and that does not provide any security level when it responds to a recommendation request.

We have modified the original ART simulator, in order to randomly generate, at the beginning of each game, the security levels of each SR and BAD agent (USR agent does not use security levels), associated with each recommendation response. These security levels are generated by a uniform distribution with values ranging in [0..1]. Both SR and BAD agents, when receive a reputation request related to an agent a, respond with a value v that reproduces the actual expertise e of a with an approximation that linearly depends on the security level l associated with v. In particular, the value v is different from e due to two errors generated by the simulator, namely err_p and err_d. err_p simulates the approximation contained in the knowledge that the requested agent has about the expertise of a. This is simulated by generating err_p using a random uniform distribution ranging in $[-\mu..\mu]$, where μ is set to 0.01. The low value chosen for μ means that the agents involved in the comparison interact with dummy agents that simulate a *training phase*, in which each dummy agent acquired a very good knowledge of the other expertise. The error err_d represents the intention of the requested agent of deceiving the requester. We have defined this error as $err_d = sign \cdot e \cdot (1-l) \cdot ((1-\delta) + \delta \cdot r)$, where $sign$ represents the sign (positive or negative) of the error, which is generated as positive in the 50% of the cases, and negative in the remaining cases, and r is a random value generated by a uniform distribution ranging in [0..1]. The maximum value of the deception error is equal to $(1-l) \cdot e$, which is inversely proportional to the security value l, while the minimum error is equal to $(1-l) \cdot (1-\delta) \cdot e$. This way, the higher the parameter δ, the smaller the minimum deception error. We have used a value $\delta = 0.05$, thus assuming that agents tend to exploit almost completely their possibility to deceive the interlocutor.



(a): Bank amount gain **(b)**: CAST vs UNO2008

Fig. 1. Experimental results

As a first campaign of experiments, we have performed the comparison in different conditions of agent population with respect to the agent expertise. In particular, this campaign was composed of 11 experiments $E1_i$, each associated with a different agent population A_i, $i = 1, \ldots, 11$. Each population A_i contains 200 agents, and it is composed of one CAST agent, one UNO2008 agent, one RRAF agent and $20 \cdot (i - 1)$ BAD agents. The remaining agents of A_i are SR agents. In other words, the agent population A_i has a percentage of inexpert agents (i.e., BAD agents) increasing with i. We have run 10 games for each experiment $E1_i$. In Fig. 1.(a), we have plotted on the vertical axis the average bank amount of each agent involved in the comparison for each experiment $E1_i$ corresponding to a different percentage of BAD agents in the agent population, where the average is computed on the 11 experiments of $E1_i$. The results clearly show that CAST is the best performing agent, and that its advantage on the other agents generally increases as the percentage of BAD agents increases. The minimum advantage on the second performer (i.e., UNO2008) is equal to 19.84% for a population having 0 BAD agents, while the maximum advantage, equal to 31%, is reached for a population containing the 80% of BAD agents. When the number of BAD agents becomes very high, the performances of all the approaches significantly worse, due to the difficulty to contact expert agents. We have performed another campaign of 11 experiments, where each experiment $E2_i$ is composed of 10 games run on a population of 200 agents, containing $(i - 1) \cdot 20$ SR agents.

In Fig. 1.(b), for each percentage of SR agents corresponding to each experiment $E2_i$, we have plotted the average percentage of advantage (computed on all the 10 games of the experiments) obtained by CAST vs UNO2008, that was almost the second performer in each game. Fig. 1.(b) highlights that, if the number of agents using the security protocol is small (i.e., less than 20% of SR agents), our approach performs significantly worse than UNO2008. Instead, if the security protocol is used enough (almost 20% of SR agents), CAST agent drastically wins, with an advantage that is about of 20%.

4 Conclusion

In this paper, we have compared CAST with other trust-based approaches in order to study the importance of the introduction of security levels. While the

results of the last ART competition rewarded UNO2008 that does not use reputation measures at all, our experiments on a set of agents larger than that used in the ART competition and including a significant percentage of deceiving agents clearly show the importance of using reputation, but only if it is combined with additional information related to the trustworthiness of the provider of reputation feedbacks. In our experiments, when a sufficient percentage of agents uses our security information, CAST already outperforms the other approaches, showing its best results in presence of an agent population characterized by very low expertise values.

Acknowledgment This work was partially funded by the Italian Ministry of Research through the PRIN Project EASE (Entity Aware Search Engines).

References

1. ART-Testbed (2011), `http://megatron.iiia.csic.es/art-testbed`
2. Garruzzo, S., Rosaci, D.: The Roles of Reliability and Reputation in Competitive Multi Agent Systems. In: Meersman, R., Dillon, T.S., Herrero, P. (eds.) OTM 2010. LNCS, vol. 6426, pp. 326–339. Springer, Heidelberg (2010)
3. Gomez, M., Sabater-Mir, J., Carbo, J., Muller, G.: Improving the art testbed, thoughts and reflections. In: Proceedings of the Workshop on Competitive Agents in the Agent Reputation and Trust Testbed at CAEPIA-2007, Salamanca, Spain, pp. 1–15. Citeseer (2007)
4. Huynh, T.D., Jennings, N.R., Shadbolt, N.R.: Certified reputation: how an agent can trust a stranger. In: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 1217–1224. ACM (2006)
5. Ismail, R., Boyd, C., Josang, A., Russell, S.: A Security Architecture for Reputation Systems. In: Bauknecht, K., Tjoa, A.M., Quirchmayr, G. (eds.) EC-Web 2003. LNCS, vol. 2738, pp. 176–185. Springer, Heidelberg (2003)
6. Jøsang, A., Golbeck, J.: Challenges for robust trust and reputation systems. In: Proceedings of the 5th International Workshop on Security and Trust Management (SMT 2009), Saint Malo, France (2009)
7. Massa, P.: A survey of trust use and modeling in current real systems. In: Trust in E-services: Technologies, Practices and Challenges. Idea Group, Inc. (2006)
8. Muñoz, V., Murillo, J.: Agent uno: Winner in the 2nd spanish art competition. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial (39), 19–27 (2008)
9. Rosaci, D.: Trust measures for competitive agents. Knowledge-Based Systems (2011)
10. Sabater-Mir, J., Paolucci, M.: On representation and aggregation of social evaluations in computational trust and reputation models. International Journal of Approximate Reasoning 46(3), 458–483 (2007)
11. Teacy, W.T.L., Huynh, T.D., Dash, R.K., Jennings, N.R., Luck, M., Patel, J.: The art of iam: The winning strategy for the 2006 competition (2007)