

This document is published in:

*Soft Computing*, (2010), 14 (8), pp. 821-831.

DOI: <http://www.dx.doi.org/10.1007/s00500-009-0470-9>

© 2009 Springer-Verlag

# An extension of a fuzzy reputation agent trust model (AFRAS) in the ART testbed

Javier Carbo, Jose Manuel Molina-Lopez

Departamento de Informatica, Universidad Carlos III de Madrid, 28911 Leganes, Madrid, Spain  
e-mail: javier.carbo@uc3m.es; e-mail: josemanuel.molina@uc3m.es  
URL: www.giaa.inf.uc3m.es

**Abstract:** With the introduction of web services, users require an automated way of determining their reliability and even their matching to personal and subjective preferences. Therefore, trust modelling of web services, managed in an autonomous way by intelligent agents, is a challenging and relevant issue. Due to the dynamic and distributed nature of web services, recommendations of web services from third parties may also play an important role to build and update automated trust models. In this context, the agent reputation and trust (ART) testbed has been used to compare trust models in three international competitions. The testbed runs locally and defines an ART appraisal domain with a simulation engine, although the trust models may be applied to any kind of automated and remote services, such as web services. Our previous works proposed an already-published trust model called AFRAS that used fuzzy sets to represent reputation of service providers and of recommenders of such services. In this paper we describe the extension required in the trust model to participate in these competitions. The extension consists of a trust strategy that applies the AFRAS trust model to the ART testbed concepts and protocols. An implementation of this extension of AFRAS trust model has participated in the (Spanish and International) 2006 ART competitions. Using this ART platform and some of the agents who participated, we executed a set of ART games to evaluate the relevance of trust strategy over trust model, and the advantage of using fuzzy representation of trust and reputation.

**Keywords:** Trust and reputation of services, Autonomous agents, Fuzzy sets

## 1 Introduction

As the amount of information in the web via html pages, wikis, blogs, and other documents grows, it gets more difficult to distinguish accurate information from inaccurate or untrustworthy information. A search engine query usually results in outdated and/or unreliable informations and the user is forced to choose manually what she/he considers the most reliable source based on her/his trust requirements. With the introduction of web services, the problem is further more complex since users have to deal with the satisfaction of their particular set of requirements for trusting recommendations about web services and web services themselves. Due to the amount and complexity of services and requirements, this task requires a more automated way of trusting each other. Two approaches to the problem of how to implement this automated computing of trust was given by Conte and Paolucci (2002): the emergent and the designed trust. Designed trust is inferred from explicit norms and social institutions observe the compliance of such norms. This view consists of a central entity that certifies the satisfaction of some given evaluation criteria. Trust is then a global property shared by all the observers. This centralized nature of trust is due to the size of these scenarios, where repeated interactions has low probability.

On the other hand, distributed approaches tackle with emergent trust, where no objective evaluation criteria are

universally accepted and the possibility of repeated meetings is not so low (Josang et al. 2007). Unlike designed models of trust, with subjective evaluations, pair-wise cooperation emerges from a spontaneous process in the form of communications about others' behaviour. In recent years, the need for computing emergent trust has become very clear. In this regard autonomous agents issuing recommendations are acknowledged as a key tool, since they would enable users to automatically select those services that match their own specific personal preferences. In other words, the solution consists of autonomous agents that form opinions about other agents based on their past services provided, as well as from references about them from third-party agents (playing the role of recommenders of such providers) in order to improve the quality of future trust decisions (Ramchurn et al. 2004). This solution can be applied to web services even if the agents are not web agents, as it is our case, but agents could be also web agents since, for instance, one of the most popular agent development platform [JADE (Bellifemine et al. 2001, 2007)] provides support for bidirectional invocation of Web services from agents, and agent services from Web service clients.

The way those opinions are formed is often called trust model, although sometimes the application of those opinions in decisions related to future interactions with other agents is also considered part of the trust model. In order to avoid confusion, we will call trust strategy to the application of the trust model.

Due to the relevance of this issue in the desired broad use of agent technology, a number of trust models and strategies have been proposed. Most of them are computational models that involve a numerical decision-making made up of utility functions, probabilities, and evaluations of past interactions (Yu and Singh 2000; Zacharia and Maes 2000; Sabater and Sierra 2001; Sen et al. 2000; Huynh et al. 2006). But from the artificial intelligence point of view, trust models embedded in agents should involve a cognitive approach (Marsh 1994): enriching the internal model of making cooperative and competitive decisions rather than enriching the exchanged reputation information. In other words, they intend to explicitly reproduce the reasoning mechanisms behind human decision-making.

Among the researchers who have proposed a socio-cognitive view of trust we could cite (Schillo and Rovatsos 2000; Castellfranchi and Falcone 1998) and (Braynov and Sandholm 2002). Schillo's model (Schillo and Rovatsos 2000) distinguish according to two types of motivation of trust: honesty and altruism of each agent. A more enriched model is from Castellfranchi and Falcone (1998) who claim that some other beliefs rather than reputation are essential to compute the amount of trust on a particular

agent: its competence (ability to act as we wish), willingness (intention to cooperate), persistence (consistency along time), and motivation (our contribution to its goals). Braynov and Sandholm (2002) highlight the relevance of modelling opponent's trust, since both agents would be interested in showing the trustworthiness of the counterpart to allocate resources efficiently.

As we have seen, very different points of view have been taken into account in trust models. These difference remark the necessity of a fair and universal (rather than ad-hoc) way to compare between proposed trust models and strategies. This is the intention of several international researchers who developed an agent reputation and trust testbed (ART) (Fullam et al. 2005a) that allow researchers to perform competitive games with objective metrics. Using such testbed, three international competitions were successfully carried out jointly with the AAMAS international Conferences of 2006, 2007, and 2008 in Hakodate (Japan), Honolulu (USA), and Estoril (Portugal), respectively. During these years the ART testbed has been widely accepted by part of the research community, while the ART-testbed members have discussed, patched, and updated the platform using the feedback from the Competitions (Team 2007b) and from the agent trust community (Team 2007a). These discussions also produced some changes in protocols (Sabater et al. 2008), and outlined new directions of work (Gomez et al. 2007).

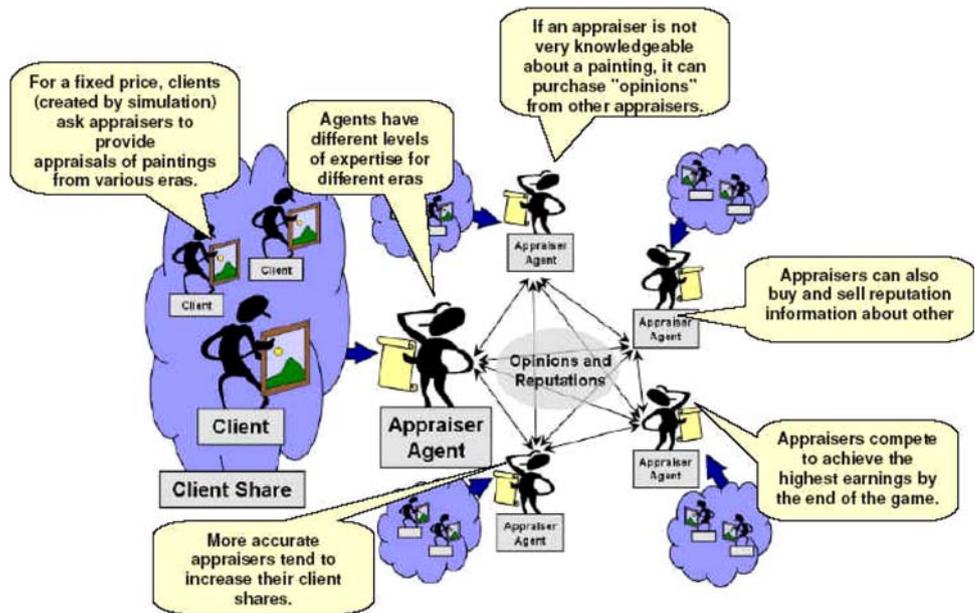
The intention of this paper is to explain how we extend the AFRAS model to be applied into ART testbed games, and to show how it would behave if some of these apparent shortcomings of ART testbed were considered. In order to achieve these goals, Sect. 2 includes an outline of the main related work: ART testbed in Subsect. 2.1 and the basics of our previously-published fuzzy trust model AFRAS in Subsect. 2.2. After that, Sect. 3 will give details about the extension applied into this AFRAS trust model in order to participate in ART competitions, which is the main contribution of this paper. Next, in Sect. 4, we will explain the games specifically designed (using ART testbed) to evaluate the relevance of trust strategy over trust model, and the advantage of using fuzzy representation of trust and reputation. Finally we will outline some conclusions and the planned future work on this issue.

## 2 Related work

### 2.1 Agent reputation and trust testbed

The ART testbed compared different trust/reputation models and strategies in an art appraisal domain. A general outline of ART interactions can be observed in Fig. 1, obtained from (Fullam et al. 2005a). In this domain, the

Fig. 1 ART domain outline



agents are the players/competitors that appraise paintings, while clients that own paintings to be appraised, instead of being autonomous agents, are emulated by the simulation engine (Fig. 2 from source (Fullam et al. 2005b))

In each timestep, a simulation engine presents each appraiser agent with paintings (generated by the simulation engine) to be appraised, paying a fixed amount (ClientFee) for each appraisal request. Very close valuation of paintings to their real value would lead to more future clients, and therefore to more earnings (called BankBalance) to win the competition. Each painting belongs to an era among a finite set of possible artistic eras while agents have different levels of expertise (ability to appraise) in each artistic eras. An agent can appraise its own paintings and may request opinions (at a fixed cost called OpinionCost) from other appraisers to get its valuation of the painting close to the real value (specially in the eras, where the agent has low expertise). An agent can also act as provider of appraisals in response to opinion (about paintings) requests from other agents. Additionally, an agent can similarly request reputation information about other appraisers (at a fixed and much lower cost than opinions called ReputationCost). According to the description given above, an agent may obtain earnings in three possible situations: when clients submit paintings to appraise to the agents; when agents sell opinions; when agents sell reputations. Agents have expenses when they buy opinions or reputations. This definition of the scenario guarantees that agents earn more BankBalance if:

1. They manage to become trusted providers of opinions and reputations,

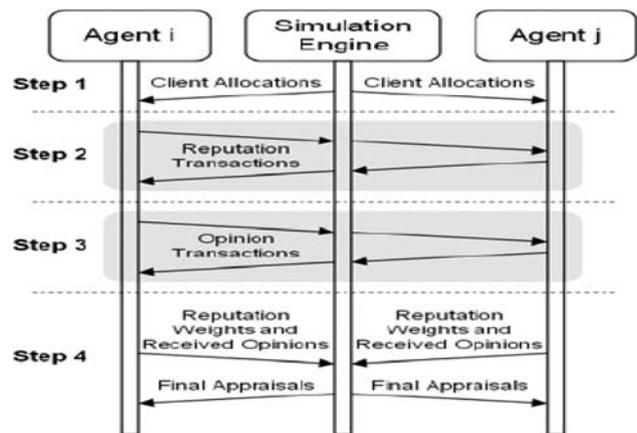


Fig. 2 The role of simulation engine in the ART testbed

2. They are able to learn as quick as possible to identify agents that can provide them with good opinions/reputations and
3. They adapt quickly to its models in cases, where the other agents change their behaviours.

Since an ART game is organized as a sequence of opinion and reputation transactions we will see below an explanation of these protocols in detail.

### 2.1.1 Opinion transaction protocol

The opinion transaction protocol suffered changes in 2008 competition<sup>1</sup> from the original specification of the testbed

<sup>1</sup> Changes for the 2008 Competition: [http://megatron.iii.csic.es/art-testbed/changes\\_2008.htm](http://megatron.iii.csic.es/art-testbed/changes_2008.htm)

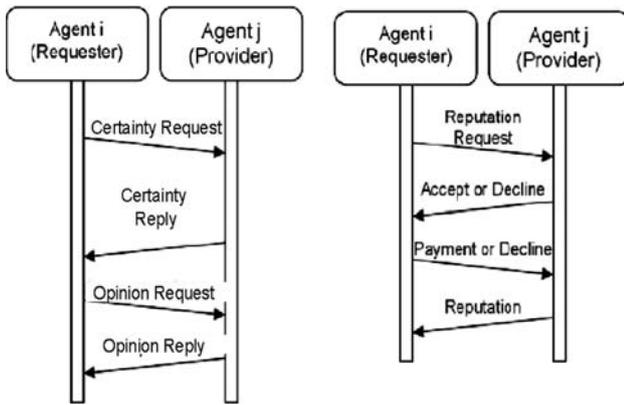


Fig. 3 ART opinion and reputation transaction protocols

(Fullam et al. 2005a). The final opinion transaction dialogue consists of two independent distinct protocols (Fig. 3).

- The first one is dedicated to ask for a certainty value of an agent in a given artistic era (where certainty refers to an assessment of agent expertise in that era). This certainty protocol begins when a requester agent sends a ‘certainty request’ message to another appraiser agent (noted as provider in advance), identifying the era of the painting to be appraised. Upon receiving the certainty request, if the potential provider is willing to provide the requested information, it responds by sending a ‘certainty reply’ message with the corresponding certainty assessment about the opinion it could provide about paintings of that era, defined as a real number between zero and one.
- The second protocol corresponds to the opinion transaction itself, where the opinion requester sends payment to the provider if it chooses to ask for an opinion from the provider.

Since direct communications between agents are not allowed, the simulation engine is in charge of forwarding such certainty and opinion messages, implementing both payments, and producing the opinions according to the corresponding expertise of the agents and to the will to provide a truthful opinion (expressed in a numerical value). However, certainty assessment is generated by the agents rather than produced by the simulation engine.

### 2.1.2 Reputation transaction protocol

Additionally, appraiser agents can exchange reputations, acting as reputation providers and requesters, following the protocol of Fig. 3. ART designers considered reputation as the information about the personal view of provider about expected expertise of other appraiser agents. A reputation

transaction begins when a requester sends a ‘reputation request’ message to a reputation provider, including the agent about which it is requesting reputation information and furthermore including the era of the expected expertise it is asking for. Then, after the provider agent has received a reputation request, it may send an accept or decline message depending on whether it is willing to provide the requested reputation (or even because it has no information at all about the trustworthiness of the given agent and era). If the provider accepted the transaction, then the requester would send the corresponding payment to the provider in order to receive the requested reputation information. Finally, the provider is not required to send its actual reputation value.

### 2.2 A fuzzy reputation model (AFRAS)

AFRAS (Carbo et al. 2003) agent adopts a socio-cognitive approach to model trust and reputation of agents. AFRAS continuously updates human-like mental attributes, expressed in fuzzy terms to adjust their progressive and smooth adaptation to the situations faced. These attributes are as egoism, sociability, shyness, and susceptibility, included in the agent reasoning. For instance, an agent would be acting socially if it answer the requests from other agents, while egoism decides whether to share or not the own knowledge or abilities with other agents, shyness stands for the proactivity of the agent asking for services, and finally susceptibility represents how much suspicious is any agent about the behaviour of others.

The internal reasoning followed by an AFRAS agent has three different levels of abstraction as it is shown in Fig. 4. They deal with world, social, and mental models. These layers are designed in such a way that each layer is bottom-up activated, and top-down executed. The former means that the higher is a layer, more complex and abstract are its

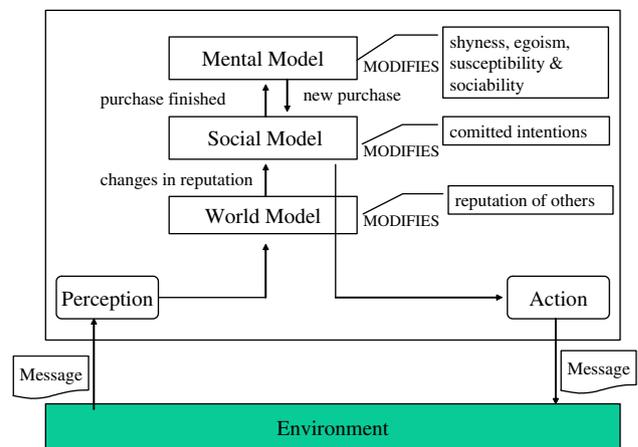


Fig. 4 Conceptual model of reasoning in AFRAS agents

competencies. The latter means that each layer uses the predicates generated by the lower layer.

The top one, called mental level, is the most interesting for this contribution since it involves the adaptation of the mental attributes of the agent that characterise its behaviour. We have called them before as shyness, egoism, susceptibility, and sociability. Its continuous adaptation affect the computations of the world model layer and the communicative decisions of the social layer and this is where trust strategy belongs to, and therefore this is the part we had to adapt to apply AFRAS model on ART games. The social model level of abstraction would activate this control layer when the results of a service evaluation are available. These mental attitudes would change according to the evolution of the results obtained by the requested services.

Furthermore, AFRAS considers trust and reputation as a fuzzy concept. This is conceptually justified because of the uncertain nature of recommendation from others, of the vague nature of evaluation of direct interactions that come from humans, and the incomplete nature of the observation of the agents (not all the actions of agents can be observed). Therefore, we represent the opinion about agents in the form of piece-wise fuzzy sets. We use an ad-hoc continuous space of values [0,100] to draw up trapezoidal fuzzy sets. We can guess that the use of fuzzy sets in trust models may facilitate the task of monitoring soft changes of behaviour. They are aggregated through a weighted mean of fuzzy sets. Aggregation of fuzzy sets is computed with Mass Assignment assumptions based on Baldwin's (1992) theory of evidence.

Additionally, reliability of trust beliefs is modelled through the fuzzy sets themselves. It is implicit in them, graphically we can interpret the gradient of the sides of a trapezium representing a fuzzy set as its reliability. In this way, a wide fuzzy set representing a given reputation represents a high degree of uncertainty over that reputation estimation, while a narrow fuzzy set implies a reliable reputation.

In a similar way, recommendations are aggregated directly with direct experiences as a weighted sum of fuzzy sets. But in this case, the weight given to each part is dependent on the trust held by the recommender.

This internal reasoning of agents has already been tested in adhoc simulations from different perspectives: the convergence of reputation (Carbo et al. 2003), the influence of benevolent recommendations (Carbo et al. 2005), and finally the dynamics of our system with a collusion of malicious providers and recommenders (Carbo et al. 2007b). It has also been considered its application to generic electronic services in the book *E-Service Intelligence—Methodologies, Technologies and Applications* (Carbo et al. 2007a) and it has been applied and compared

with the web service provided by MovieLens Website in the *International Journal of Web Engineering and Technology* (Carbo and Molina 2004).

### 3 An extension of AFRAS to be applied in ART testbed

Any agent addressing trust problems has to implement decisions on how to use trust and reputation information (trust strategy), but AFRAS just comes up with updating such information, not applying it. This means that we have to extend AFRAS model, and in particular for the ART domain, we have to implement the following decisions:

1. Whether or not asking for reputation of a service provider (reputation request in ART terms) to other given agent,
2. Whether or not answering a service (reputation/opinion) request, and
3. Whether or not asking for a service (requesting an opinion).
4. How much time (economic value) our agent should spend building opinions to the 'ask for opinions' request about the paintings of the other agents
5. How much time (economic value) our agent should spend building the appraisals of the own paintings

In the design of AFRAS internal reasoning schema, these decisions should be taken in the world model layer according to the particular state of the mental characterisation of the agent, and they do not need to be changed. However, the updating of egoism, sociability, and shyness, that takes place in the mental layer has to be implemented in order to act in ART games. This is a flaw of our previously-published and previously-tested AFRAS trust model since it exclusively updates trust beliefs with direct and indirect sources (noted in ART terms as opinions and reputations) and susceptibility attribute.

Therefore, in this paper we design and implement a trust strategy for extending AFRAS trust model to be used in ART testbed. This extension consists of an ART ad hoc instantiation of some of the above-mentioned human-like attributes of AFRAS model: Shyness, egoism and sociability. Susceptibility is excluded to this extension since it belongs to the trust model (it only influences the adaptation of trust updates with new information). It has to be noticed that then the level of success of last prediction involved in susceptibility updating remains being computed as a similarity function of the overlapping between two fuzzy sets that represents the prediction and the result obtained. But since the result obtained is not anymore a fuzzy set and although AFRAS trust model assumes that agents are exchanging trust values in the form of piece-wise fuzzy sets, this is not allowed in ART testbed (since exchanged

data in ART communications are limited to float values between zero and one). So we have overcome this limitation transforming a float value in a very thin fuzzy triangle centered in such float value of trust. By this way the updating of susceptibility has not changed.

This trust strategy has to take into account the economic incomes and expenses of ART games: the generation of appraisals for us (rewarded with a fixed ClientFee for each painting), the investment in opinion for others and finally providing and requesting opinions and reputations. Requesting opinions to other appraiser agents has a cost noted as OpinionCost (fixed in the game design definition), and providing them has an equivalent income. On the other hand, requesting/providing reputation has a cost/income of ReputationCost. Since ReputationCost value (defined in ART competitions as 1) is much lower than the OpinionCost (10) and ClientFee (100), AFRAS agent always ask for reputation to every agent in the system (there were a few number of agents in the games of past competitions, this is one of the flaws of them (Gomez et al. 2007)). Additionally we want AFRAS agent to always accept all the relevant incomes (ask-for-opinions rather than ask-for-reputations). Therefore, our agent is always completely sociable, this means that one of our mental attributes, Sociability, is never adapted or updated, it has a maximum value along all the game. We chose these very simple decisions in order to avoid much more noise in the results, since we want to test just the convenience of AFRAS (at least as much as we can) instead of the convenience of a particular economic strategy.

So it still lacks the third of the economic decisions included in the trust strategy: how many agents should ask for opinions about its assigned paintings (each request costs an OpinionCost value). In the AFRAS reasoning this decision had to be taken using the Shyness mental attribute of the agent, and we now have to decide how to update it. We will update Shyness attribute through the following guidelines:

- In the first timestep AFRAS agent would be not shy at all since it requests all opinions but the worst one, considering the worst one according to the certainties received from the other agents in the game (agents advertise its expertise in generating opinions to the requesting agents before the opinion request is confirmed).
- In the next timestep AFRAS agent asks for opinion of every *interesting* agent, where *Interesting* is defined below.

Interesting agents are the first ones in an ordered list by their reputation. We assume the statement that number of interesting agents has to be decreased as much as the

overall level of success of the agent estimating the real value of the paintings is increased. This level of real success is represented in *Remembrance* variable from the last increment of the number of assigned paintings (# of *appraisals*). This number is proportional to the quality of the agent results in the past iteration, and it is decided for each agent by the internal computations of the ART system. We can also consider *Remembrance* as the general confidence of the agent on all of its own beliefs in a particular timestep (i).

$$\text{Remembrance}_i = \frac{\# \text{of appraisals}_i}{\# \text{of appraisals}_{i-1}} \quad (1)$$

Then, the computation of the number of interesting agents is obtained from the defuzzification of the union of the next fuzzy sets corresponding to the *Shyness* of the agent in a timestep i:

$$\begin{aligned} \text{FzShyness}_i = & \text{union}\{\text{FzShyness}_{i-1}\} \\ & \{[0, 0][\# \text{agents} - 1 - \text{Remembrance}_i, 1] \\ & [\# \text{agents}, 0]\} \end{aligned}$$

$$\# \text{of interesting agents}_i = \text{defuzzify}\{\text{FzShyness}_i\}$$

where there are several ways to implement a defuzzification function (centroid of area, bisector of area, mean value of the maximum, etc.). We chose a centroid defuzzification method. The next (forth) economic decision of our trust strategy is: how much time (economic value) our agent should spend building opinions to the ‘ask for opinions’ request about the paintings of the other agents. In the AFRAS model this decision was taken using the *Egoism* mental attribute of the agent. In the corresponding extension of AFRAS this economic decision is updated through the following guidelines:

- In the first timestep our agent is not egoist at all since it spends much time building such answers (opinions for the other agents). This implies a great economic investment to build a high initial reputation of our agent for the others.
- And gradually time spent building opinions (economic inversion) depends upon the Remembrance value and upon the economic efficiency (relative benefits per appraisal, noted *OpportunityCost*) obtained by the agent.

Opportunity costs value is computed from the potential benefit we could have obtained without any expenses asking for opinions:

$$\text{OpportunityCost}_i = 1 - \frac{(\text{BankBalanceInc}_i)}{\# \text{of Appraisals}_i * \text{ClientFee}}; \quad (2)$$

where *ClientFee* is the amount an agent receives for each appraisal (fixed in advance by ART system) and the

relative difference between the *BankBalance* in two sequential timesteps is *BankBalanceInc*. Now we take into account the last level of success and the efficiency of our agent to compute the egoism value of timestep  $i$ :

$$\text{Egoism} = \frac{\text{ClientFee}}{\text{OpinionCost}} - \text{Remembrance}_i - \text{OpportunityCost}_i, \quad (3)$$

where *OpinionCost* is the amount that any agent has to pay to other agents for obtaining an opinion about a painting (also fixed in advance by ART system). And this crisp value is fuzzified as we did with shyness, and then, we apply a union operation between the last fuzzy egoism and the previous ones:

$$\text{FzEgoism}_i = \text{union}\{\text{FzEgoism}_{i-1}\} \left\{ [0, 0][\text{Egoism}, 1] \left[ \frac{\text{ClientFee}}{\text{OpinionCost}}, 0 \right] \right\}$$

This fuzzy set will be used to decide the quantity of time to be invested in the opinion generation of timestep  $i$  (*ExternalInvest*):

$$\text{ExternalInvest}_i = \text{OpinionCost} * \text{Defuzzify}(\text{FzEgoism}_i) \quad (4)$$

Here again the defuzzification method is computed from the centroid of the area. The last one (fifth) of the economic decisions that belongs to our trust strategy is how much time (economic value) our agent should spend building the appraisals of the own paintings. Our approach follows the general balancing rule that implies investing more in the eras where more is needed. We will need to invest more in the eras where we lack our expertise (*OwnExpertise*) plus the expected expertise provided by other agents (*ExpectedExpertise*). Therefore *Necessity* is defined from the total accumulated expertise about the corresponding painting.

$$\text{Necessity}_j = \frac{\#\text{Agents} - \text{OwnExpertise}_j - \text{ExpectedExpertise}_j}{\#\text{Agents}} \quad (5)$$

Then the time invested building our opinions about our own painting  $j$  (*InternalInvest*) is computed from:

$$\text{InternalInvest}_j = \frac{\text{ClientFee}}{\text{OpinionCost}} * (1 + \text{Necessity}_j) \quad (6)$$

#### 4 AFRAS tests in ART games

An implementation of AFRAS trust model was presented to 2006 international ART competition. The results

**Table 1** 2006 ART competition results

Rank	Average score	Agent name	Team representative	Team affiliation
1	132321	IAM	Huynh	Univ. Southampton
2	123008	joey	Thiel	Univ. Nebraska
3	121443	neil	Weng	Nanyang Tech. Univ.
4	116464	frost	Kafali	Bogazici Univ.
...				
12	92796	AFRAS	Carbo	Univ. Carlos III
...				

obtained by this agent in such competition were not impressive (12th position over 17 registered agents), as Table 1 shows.

Due to these disappointing results we could argue whether fuzzy representation makes sense when we use reputation information to infer trust or not. However, it is not clear how much benefits come from strategic trust decisions (and their particular instantiation explained here) rather than the classical trust decisions (just updating information with direct/indirect experiences), and furthermore and how much come from the fuzzy representation. Following some of the current discussion about how to obtain more detailed conclusions for the next ART competitions of ART testbed (Gomez et al. 2007), we have designed and run specific ART games to give an approximated response to both questions: relevance of trust strategy and of fuzzy representation.

From our point of view, one of the most relevant design failures of past ART competitions is the lack of necessity of reputation transactions (in fact, the winner, IAM (LukeTeacy et al. 2007), did not use them at all). ART developers tried to solve this problem (from 2006 to 2007 competition) increasing the scalability of ART testbed (Team 2007b).

According to both documents (Gomez et al. 2007; Team 2007b) we chose to apply a very different initial setup of the games from the competitive ones, that we think would overcome partially the problem of uselessness of the reputation transactions. These changes are:

- We set the number of eras to 10 rather than 5, and the number of initial appraisals to 5 rather than 20. By these values, it is harder to acquire enough direct knowledge from opinions (since there are more eras to know about, and less initial paintings to evaluate), and then reputation is much required than before.
- We set the opinion cost to 50 rather than 10. Then reputation transactions is worth, since opinion transactions are much more expensive.

**Table 2** Publicly available 2007 ART registered agents

Agent name	Team representative	Team affiliation
IAM	Luke Teacy	Univ. Southampton
jam	Anil Gursel	Univ. Tulsa
zecariocales	Andrew Diniz	Pontificia Univ. Rio de Janeiro
spartan	Nicolas Hormazabal	Univ. of Girona
artgente	Mikalai Sabel	Univ. Trento
uno	Victor Muoz	Univ. of Girona
reneil	Jianshu Weng	Nanyang Tech. Univ.
marmota	Javier Murillo	Univ. of Girona

#### 4.1 ART games proving the suitability of fuzzy representation

Since the main point of our proposal is the use of a fuzzy representation of trust and of the agent’s mental characterization, we would like to observe the influence of the fuzziness much more than the overall comparison with other trust models held in ART competitions. Therefore, we thought that a way to test such influence is to observe the relative differences of behaviour with a crisp clone of our AFRAS agent when this new agent (we will call it ACRAS, A Crisp Reputation Agent System) faces the same games of the international ART competitions. Although AFRAS just participated in 2006 competition (under the name of Fzjcarbo), we used the finalist agents of 2007 competition and the enhanced version of the testbed of 2007. We did it because the same university (Southampton) won both years (with IAM and IAM2), and due to the enhancements of the testbed that did not avoid the participation of the agents from 2006 as our Fzjcarbo agent. Unfortunately 2006 and 2007 agents do not run in the last version of 2008. The compiled code of all the participants in the competitions, and the testbed itself are publicly available in the ART website.<sup>2</sup> The 2007 top participants were those of Table 2.

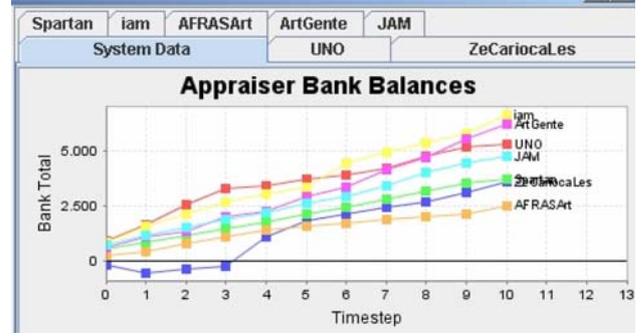
Therefore to build the crisp equivalent to AFRAS, we had to transform each fuzzy variable and each fuzzy computation into an equivalent crisp one. So then the computation of shyness from:

$$\text{FzShyness}_i = \text{union}\{\text{FzShyness}_{i-1}\} \\ \{[0, 0][\#\text{agents} - 1 - \text{Remembrance}_i, 1] \\ [\#\text{agents}, 0]\}$$

is defuzzified as follows:

$$\text{Shyness}_i = (\text{Shyness}_{i-1} + \#\text{agents} - 1 \\ - \text{Remembrance}_i)/2 \quad (7)$$

The computation of egoism from:

**Fig. 5** Sample game with AFRAS and the finalists of 2007 ART competition

$$\text{FzEgoism}_i = \text{union}\{\text{FzEgoism}_{i-1}\}$$

$$\left\{ [0, 0][\text{Egoism}, 1] \left[ \frac{\text{ClientFee}}{\text{OpinionCost}} \right] \right\}$$

...goes into:

$$\text{Egoism}_i = (\text{Egoism}_{i-1} + \text{Egoism})/2 \quad (8)$$

where the weighted sums of fuzzy sets corresponding to the reputation aggregation, and trust updating of providers and recommenders (from opinion responses, reputation responses, and real value of paintings) are straightly converted into weighted sums of float trust values. Furthermore, all the similarity computations obtained from the overlapping of fuzzy sets are converted into a simple difference between two float values.

Once the crisp version of AFRAS was ready, we run two series of the same ART games (facing the 2007 competition finalists) of 10 timesteps with both alternative agents, the fuzzy original one (AFRAS) and the crisp clone (ACRAS). In the Fig. 5 we can observe a capture from the ART interface with the corresponding Bank Balances of a sample of these games, where AFRASArt obtained the last position with 2503 and the winner (IAM) obtained 6523.

We obtained the next average result (the bank balance means and their standard deviations) of 50 games in Table 3.

From these very close results between AFRAS and ACRAS facing the best agents of 2007 competition, we compute a comparison of means (*t* test) in order to state the significance of hypothesis  $H_0: \mu_1 = \mu_2$  since the difference in mean shows just a slightly better results for AFRAS.

**Table 3** 20 ART games results

Agent name	Earnings mean	Earnings Std Dev
ACRASArt	2704.1	800.87
AFRASArt	2811.85	772

<sup>2</sup> <http://www.art-testbed.net>

Therefore, considering a (approximately) normal distribution centered in zero for  $H_0 : d_{\bar{x}} = 0$ , we compute:

$$d_{\bar{x}} = \bar{x}_1 - \bar{x}_2 = 2811.85 - 2704.1 = 107.75 \quad (9)$$

where standard deviation of  $d_{\bar{x}}$  is:

$$s_{d_{\bar{x}}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (10)$$

where it results:

$$s_{d_{\bar{x}}} = \sqrt{\frac{772^2}{50} + \frac{800.87^2}{50}} = 157.31 \quad (11)$$

And then the condition to accept  $H_0$  would be:

$$d_{\bar{x}} \leq Z * s_{d_{\bar{x}}} \quad (12)$$

where considering a significance of 99% ( $\alpha = 0.01$ ), and the corresponding  $Z_{\alpha}$  of 2.58, we can find out:

$$107.75 \leq 2.58 * 157.31 = 405.87 \quad (13)$$

And therefore we cannot state a relevant difference between AFRAS and ACRAS Bank Balance means.

#### 4.2 ART games proving the relevance of trust strategy over trust model

One of the reasons we found for the ART competition results was that AFRAS agent was defined as a trust model rather than a trust strategy. The focus of AFRAS (and our general point of view about trust models) is to find the way to form the right valuation about others from direct experiences and indirect references rather than finding the right way to use them in the most efficient way (trust strategy). Since ART competition design did not distinguish between them, we thought in a way to reduce the influence of trust strategy over trust model decisions through the game setup and through the implementation of the agent participants. By both ways we would not change a line in the code of ART testbed itself.

In ART competitions, agents acted jointly as provider and client roles, and we think that these multiple roles of agents may be disturbing ART results to show which participant was the best one just modelling the trust of other agents. One way to reduce the influence of trust strategy over trust model is to avoid agent participants acting as providers of opinions. But since ART code includes this possibility of agents, we will isolate both roles by the behaviour of the agents participants. Therefore, in our ART games we implement just pair-wise comparisons. They include a variant of our AFRAS agent of the 2006 competition adapted to 2007 games and it does not act as provider: AFRASArt (it provides a zero opinion certainty to every opinion certainty request), and additionally we

include the winner of the competition, IAM (LukeTeacy et al. 2007) from University of Southampton, just as it was in 2007 competition. Since we cannot modify the source code of IAM agent, the only way to avoid IAM acting as provider is to avoid any agent in the game to ask him for opinions. So we also add an ad hoc modification into our AFRASArt agent in order to specifically not asking IAM agent for opinions. And finally, all the other agents that fill the game (we will call them 'predefined agents'), are non participants in the comparison, they are designed to provide opinions to IAM and AFRASArt agents and to not asking both of them for opinions. Consequently AFRASArt and IAM do not act as providers of opinions to others in our games, so their earnings are not obtained providing opinions to other agents.

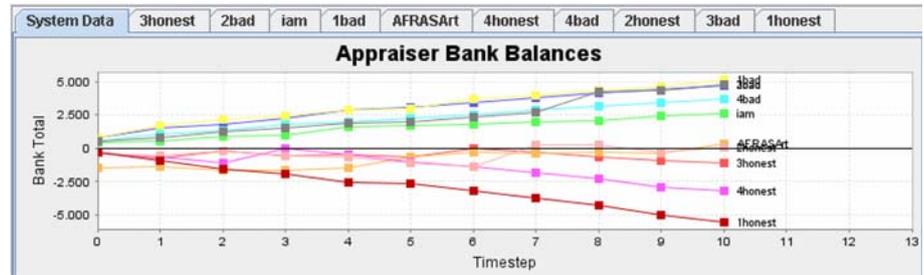
However, we are not avoiding completely the provider role of IAM agent since it still invests in their own appraisals (it was designed to do it, and we cannot access to IAM agent source code, but we know how much they invest in them from (LukeTeacy et al. 2007)). So, we decide that AFRASArt also may invest in own appraisals, and that a way to minimise the relevance of this investment on the own appraisals of IAM is to increase very much the investment of some of the predefined agents (the ones we will called 'honest') when they generate opinions for anyone. So the possibility of counting on much better evaluations of paintings from others would reduce the final relevance of this acting as providers of IAM and AFRASArt agents.

Through the use of two opposite types of predefined agents, bad and honest agents, we intend to push IAM and AFRASArt agents to filter out bad agents when asking for opinions, while to promote asking for opinions to honest agents. Since both types compute and exchange internal true trust values in reputation transactions, then reputation transactions are more useful than they were in past ART competitions, where many participant agents share false information about third parties in reputation transactions.

Specifically these two types of predefined agents in our games behave in the following ways:

- Honest agents, who provide very high-quality opinions (investing twice the opinion cost), and ask for opinions to all the other agents (but IAM and IMM agents), computing trust as an average sum of opinions provided. They do not ask for reputation of third parties, but they provide honestly their trust values when they are asked for the reputation of third parties.
- Bad agents, who are similar to the honest ones, but they provide very low-quality opinions (investing the minimum value allowed). As they provide honest trust values in the reputation transactions, they are considered just bad agents (as if they would have very low expertise

**Fig. 6** Sample game with IAM, AFRAS, 4 honest, and 4 bad agents



**Table 4** 20 ART games results

Agent name	Earnings means	Std Dev
IAM	2391.6	338.13
AFRAS	1635.75	687.65

values in all the eras) rather than malicious, since they compute (and exchange) trust values of third parties (as reputation responses) in the same honest way that the so called honest agents did.

Below we show a figure (Fig. 6) from the ART interface of a sample game with IAM, AFRAS, 4 honest, and 4 bad agents.

Since a single game is not significant, Table 4 shows the bank balance means and standard deviations of 50 games we run with 4 honest agents, 4 bad agents jointly with IAM and AFRAS agents.

Computing again  $t$  test on the difference of these means, for a (approximately) normal distribution centered in zero for  $H_0 : d_{\bar{x}} = 0$ , we find out:

$$d_{\bar{x}} = \bar{x}_1 - \bar{x}_2 = 2391.6 - 1635.75 = 755.85 \quad (14)$$

$$s_{d\bar{x}} = \sqrt{\frac{338.13^2}{50} + \frac{687.65^2}{50}} = 108.37 \quad (15)$$

And therefore the acceptance of  $H_0$ , considering a significance of 99% ( $\alpha = 0.01$ ), and the corresponding  $Z_{\alpha}$  of 2.58, it results:

$$755.85 > 2.58 * 108.37 = 279.59 \quad (16)$$

As these data show, the bank balances of IAM agent and AFRAS agent are different enough, although their means are slightly more close than in the official competition (28% of relative distance between them rather than 30%).

We can guess that the authors of IAM could argue that this is not a fair comparison since we have modified the environment conditions of the game (although using the same ART testbed code). Obviously if the authors of IAM agent have previously known the setup of this different kind of game, then they might implement changes in IAM agent to include an ad hoc trust strategy to the new setup. But our point was not about claiming anything about its winning position in 2007 competition.

Instead of it, these 50 games intend to point out that avoiding part of the economic incomes (those that come from the role of providers) and that promoting reputation utility, ART testbed games may be (at least a bit) more accurate. And particularly from our view, these games tried to show whether AFRAS principles may be used as a reasonably good trust model or as a reasonably good trust strategy. But these results are not decisive for any of these clauses. We just saw that AFRAS model and strategy share both in similar ways the responsibility of the disappointing results of AFRAS in the 2006 ART competition. This conclusion could be even considered a surprise since an analogous study over a participant agent that used IMM filter as working core showed relevant differences between the results of its trust model over its trust strategy (Carbo et al. 2008).

## 5 Conclusions

Due to the increasing number of Web services, there is a need of an automated way of inferring their reliability (trust on them). Agent technology seems to be the right approach when trust depends on peer to peer opinions. In this line, the outstanding Agent Reputation and Trust (ART) initiative had a significant acceptance in providing a tool for computing fair comparisons of trust models. However, some shortcomings of ART testbed view were pointed out. The inclusion of our fuzzy-based AFRAS trust model into an ART participant agent was therefore justified. Although the original trust model was widely published and tested some time ago (Carbo et al. 2003), and its application to the ART testbed meant significant extensions and modifications that were described in this paper. This new agent that includes the extension of AFRAS model was tested in the ART games corresponding to the Spanish and the international competitions held in 2006.

The intention of overcoming the disappointing results obtained by the AFRAS extension to the ART domain rules in these competitions, drove us to design two different ART games to evaluate in some way the positive influence of a fuzzy representation in this context, and the strength of the dependence of the results on the trust strategy rather

than the trust model (which was supposed to be the strong point of AFRAS). Unfortunately the results of these games did not provide us enough encouragement for both statements. In spite of that, we think that our fuzzy approximation to the trust problem still makes sense and we hope that future enhancements of the testbed platform and rules would contribute to a better clarification on the results that ART games are measuring.

**Acknowledgments** This work was supported in part by projects CICYT TIN2008-06742-C02-02-TSI, CICYT-TEC2008-06732-C02-02-TEC, SINPROB, CAM MADRINET S-505-TIC-0255 and DPS2008-07029-C02-02.

## References

- Baldwin J (1992) A calculus for mass assignments in evidential reasoning. In: Fedrizzi M, Kacprzyk J, Yager R (eds) *Advances in the Dempster-Shafer Theory of Evidence*. Wiley, New York, US, pp 513–531
- Bellifemine F, Poggi A, Rimassa G (2001) Developing multi-agent systems with a FIPA-compliant agent framework. *Softw Pract Exp* 31:103–128
- Bellifemine FL, Caire G, Greenwood D (2007) *Developing multi-agent systems with JADE (Wiley Series in Agent Technology)*. Wiley
- Braynov S, Sandholm T (2002) Trust revelation in multiagent interaction. In: *Workshop on the philosophy and design of socially adept technologies*. Minneapolis, pp 57–60
- Carbo J, Molina J (2004) Agent-based Collaborative Filtering based on Fuzzy Recommendations. *Int J Web Eng Technol* 1(4):414–426
- Carbo J, Molina J, Davila J (2003) Trust management through fuzzy reputation. *Int J Coop Inf Syst* 12(1):135–155
- Carbo J, Molina J, Davila J (2005) Fuzzy referral based cooperation in social networks of agents. *Artif Intell Commun* 18(1):1–13
- Carbo J, Garcia J, Molina J (2007a) Trust and Reputation in e-services: concepts, models and applications. In: *E-Service intelligence—methodologies, technologies and applications*. Springer Verlag
- Carbo J, Molina J, Davila J (2007b) Avoiding malicious agents using fuzzy recommendations. *J Org Comput Electron Commerce* 17(1):101–117
- Carbo J, Garcia J, Molina JM (2008) Noise detection in agent reputation models using IMM filtering. In: *Trust in agent societies*. Lecture Notes in Computer Science 5396, pp 25–42
- Castellfranchi C, Falcone R (1998) Principles of trust for multiagent systems: cognitive anatomy, social importance and quantification. In: *Third international conference on multi-agent systems*. pp 72–79
- Conte R, Paolucci M (2002) *Reputation in artificial societies*. Kluwer Academic Publishers
- Fullam K, Klos T, Muller G, Sabater J, Schlosser A, Topol Z, Barber KS, Rosenschein J, Vercouter L, Voss M (2005a) A Specification of the agent reputation and trust (ART) testbed. In: *4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2005)*. pp 512–518
- Fullam K, Klos T, Muller G, Sabater J, Topol Z, Barber KS, Rosenschein J, Vercouter L (2005b) The Agent Reputation and Trust (ART) Testbed Architecture. In: *Workshop on Trust in Agent Societies at The Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2005)*. pp 50–62
- Gomez M, Sabater-Mir J, Carbo J, Muller G (2007) Improving the ART testbed, thoughts and reflections. In: *Proceedings of 12th CAEPIA conference*, pp 1–15
- Huynh TD, Jennings NR, Shadbolt NR (2006) An integrated trust and reputation model for open multi-agent systems. *Auton Agent Multi Agent Syst* 13(2):119–154
- Josang A, Ismail R, Boyd C (2007) A survey of trust and reputation systems for online service provision. *Decis Support Syst* 43(2):618–644
- LukeTeacy W, Huynh T, Dash R, Jennings N, Patel J, Luck M (2007) The ART of IAM: the winning strategy for the 2006 competition. In: *Proceedings of trust in agent societies WS Procs., AAMAS 2007*
- M. Schillo PF, Rovatsos M (2000) Using trust for detecting deceitful agents in artificial societies. *Appl Artif Intell* 14(8):825–849. Special Issue on Trust, Deception and Fraud in Agent Societies
- Marsh S (1994) Trust in distributed artificial intelligence. In: Castellfranchi, Werner (eds) *Lecture notes in artificial intelligence* 830. Springer Verlag, pp 94–112
- Ramchurn SD, Huynh T, Jennings N (2004) Trust in multiagent systems. *Knowl Eng Rev* 19(1):1–25
- Sabater J, Gomez M, Muller G, Carbo J (2008) Changes for the 2008 Competition
- Sabater J, Sierra C (2001) Regret: a reputation model for gregarious societies. In: *Fourth workshop on deception, fraud and trust in agent societies*. Montreal, Canada, pp 61–69
- Sen S, Biswas A, Debnath S (2000) Believing others: pros and cons. In: *Proceedings of the 4th international conference on multi-Agent systems*. Boston, MA, pp 279–285
- Team A (2007a) ART Testbed Discussion Board
- Team A (2007b) Testbed-related discussion notes
- Yu B, Singh M (2000) A social mechanism for reputation management in electronic communities. *Lect Notes Comput Sci* 1860:154–165
- Zacharia G, Maes P (2000) Trust management through reputation mechanisms. *Appl Artif Intell* 14:881–907