An Anticipatory Trust Model for Open Distributed Systems

Mario Gómez, Javier Carbó, Clara Benac-Earle
{mgomez1,jcarbo,cbenac}@inf.uc3m.es

Group of Applied Artificial Inteligence
Carlos III University of Madrid
Av. Universidad 30, 28911, Leganés (Spain)

Abstract. Competitive distributed systems pose a challenge to trust modelling due to the dynamic nature of these systems (eg. electronic auctions) and the unreliability of self-interested agents. We propose a trust model which does not assume a concrete cognitive model for other agents an agent may interact with, but uses the discrepancy between the information provided by other agents and its own experience in order to anticipate their actions. By anticipating the behavior of other agents, and agent is able to adapt more effectively to changes in the environment for its own benefit.

1 Introduction and motivation

Although there are different definitions [7], we can state that trust is an abstract property applied to others that helps to reduce the complexity of decisions. Trust is a universal concept that plays a very important role in social organizations as a mechanism of social control. Therefore, modelling trust in open distributed systems such as agent systems becomes a critical issue since their offline and large-scale nature weaken the social control of direct interactions. For this reason, the agent research community has been very interested in this issue.

Often, there are objective and universal criteria to evaluate the quality of interactions (products/services provided by them). In this case, trust can be inferred from certificates issued by third parties that verify such objective criteria. Unfortunately, when a set of universal objective evaluation criteria is not available, this subjective and local trust will not be easily asserted. There are several application domains where interpersonal communications are the main source of trust due to the subjective nature of the evaluation criteria (books, films, web pages, leisure activities, consulting services, technical assistance, etc.).

Although there are several ways to infer trust, numerous studies have shown that in real life one of the most effective channels to avoid deceptions is through reputation-based mechanisms [2]. Usually, the group of people with good reputation (collaborators, colleagues and friends) that cooperates with a particular person to improve the quality of decisions forms an informal social network [13]. In this context, trust and reputation are strongly linked. Several trust models have been proposed ([1],[10], [15], [12], [14], [6], [3], [5]). Here we give a brief overview of some of the most relevant ones.

Two of the most cited reputation models are SPORAS and HISTOS [15]. SPORAS is inspired in the foundations of the chess players evaluation system called ELOS. In this model, trusted agents with very high reputation experience much smaller changes in reputation than agents with low reputation. SPORAS also computes the reliability of agents' reputation using the standard deviation of such measure. HISTOS is designed to complement SPORAS by including witness information as a second source of reputation.

The REGRET model [12] takes into account three types of information sources: system, neighborhood and witness reputations. REGRET includes a measure of the social credibility of the agent and a measure of the credibility of the information in the computation of witness reputation. The first of them is computed from the social relations shared between both agents. The second measure, information credibility, is computed from the difference between the recommendation and what the agent experienced by itself. REGRET establishes an intimacy level for interactions to measure the confidence on the beliefs induced by those interactions.

The Singh and Yu trust model [14] uses Dempster-Shafer theory of evidence to aggregate recommendations from different witnesses. The main characteristic of this model is the relative importance of fails over success. It assumes that deceptions (valued negatively) cause stronger impressions than satisfactions (valued positively). It then applies different gradients to the curves of gaining/losing reputation in order to lose reputation easily, while it is hard to acquire it.

From the artificial intelligence point of view, computational models of trust embedded in agents involve a cognitive approach [10]: modelling opponents to support cooperations and competitions. In contrast, computational models of reputation in agents involve a numerical approach, made up of utility functions, probabilities and evaluations of past interactions. The combination of both computational models intends to reproduce the reasoning mechanisms behind human decision-making. In this paper we present a trust modelling framework that combines both views, since it assumes the cognitive stance, but uses a numerical approach.

Other researchers have proposed a socio-cognitive view of trust [6], [3]. As an example, the socio-cognitive approach from Carbo et al. [5] supports the fuzzy nature of the reputation concept itself. Additionally it also includes other beliefs in the AFRAS trust model that intend to represent an emotive characterization of agents (shyness, egoism, susceptibility). It also includes a global belief, *remembrance*, to represent the general confidence of the agent on its own beliefs.

Many existing approaches to trust modelling have paid little attention to a crucial feature of autonomous agents: their capacity to be pro-active rather than just reactive, i.e., their ability to deal with the future by mental representations or specific forms of learning. For guiding and orienting a future action, a representation of the future, and more precisely, a representation of future effects and of intermediate results of the action, is needed [9]. Anticipatory behavior is an interdisciplinary topic attracting attention from computer scientists, psychologists, philosophers, neuroscientists, and biologists [4]. Anticipation can be seen as mechanism for devising hypotheses that make predictions about future events, conducting experiments to corroborate them and subsequently using the knowledge gained to perform useful behaviors. Anticipatory principles are interesting in the context of trust and reputation modelling because they define a continuing process of discovery and refinement that would allow an agent to adapt quicker to dynamic environments.

In this paper we present the components of an anticipatory model to handle computational trust in dynamic distributed environments. The paper is organized as follows: section §2 describes the different components of trust and presents a synthetic definition of trust as an aggregation of its components, §3 describes some experiments to test our model using the ART Testbed, and finally, §4 sums ups the contributions of our work.

2 The Anticipatory Trust Model

Typically, a trust model considers two main sources of information to estimate trust: direct experience, sometimes referred to as direct trust or interaction trust, and recommendations, often called witness-information or "word of mouth". In our model we keep this distinction between direct experience and recommendations, but in addition, we distinguish between the recommendations about third party agents and the recommendations provided by an agent about itself, what we call *advertisements*. All in all, our model builds trust upon three components, namely: Direct Trust (DT), Advertisements-based Trust (AT), and Recommendations-based Trust (RT).

In order to adapt quicker to the dynamic and uncertain nature of an open environment, an agent can anticipate or have expectations (not necessarily rational) about the possible consequences of its actions, therefore, we distinguish between the historic components of trust, based on past information only, and the anticipatory components.

In our model, only the Advertisements-based Trust and the Recommendations-based trust are anticipatory, while trust by direct experience is purely an historic belief. To simplify the dynamics of a multi-agent system, we use a discrete time model made up of time steps. A time step represents the minimal time period an agent requires to take decisions, act, and perceive the result of its actions. We use t to denote a particular time step in the past, T for the current time step, $T+1$ for the next time step, and ΣT for an aggregation of historic beliefs until time step T.

To handle uncertainty and ignorance, we use two dimensions to represent the confidence on a belief, namely: *intimacy*, and *predictability*. Intimacy is a measure of confidence based on the number of data (or interactions) used to calculate a belief, while predictability is a measure of confidence based on the dispersion or variability of the data. In our model, all the components of trust have attached a measure of confidence made up of intimacy and predictability. In addition, we propose the use of t-norms for combining intimacy and predictability into a single confidence value, and t-conorms for calculating the confidence coming from several sources of information.

Direct Trust($DT_{ij}^{\Sigma T}$): assesses the Quality of Service i provided from agent j until time step T inclusive.

$$DT_{ij}^{\Sigma T} = \frac{\sum_{t=0}^{T} \varphi(t,T) pDT_{ij}^{t}}{\sum_{t=0}^{T} \varphi(t,T)} \tag{1}$$

where $pDT_{ij}^{t} : \mathbb{R} \rightarrow [0,1]$ is the partial Direct Trust obtained for agent j and service i in time step t, and $\varphi(T,t) : \mathbb{N} \rightarrow [0,1]$ is a forgetting function used to weight each partial belief according to its age (number of time steps since a belief was obtained, T-t).

Direct Trust Confidence($DTC_{ij}^{\Sigma T}$): assesses the reliance of Direct Trust as an estimator of the Quality of Service i provided by agent j.

$$DTC_{ij}^{\Sigma T} = ITM_{ij}^{DT} \otimes (1 - \upsilon_{dt}(pDT_{ij}^{t})) \tag{2}$$

where $ITM_{ij}^{DT} \in [0,1]$ is the intimacy level for DT ([12]), a growing function in [0,1] over the number of pDTs used to compute DT, $\upsilon_{dt} \in [0,1]$ is a measure of the variability of pDT_{ij}^{t}, and $\otimes$ is a *T-norm* operator.

Advertisements-based Trust(AT_{ij}^{T+1}): assesses the Quality of Service i expected from agent j in the next time step $(T+1)$, based on advertisements.

$$AT_{ij}^{T+1} = \left\{ \begin{array}{cc} 1 & pAT_{ij}^{T+1} + \Delta AT_{ij}^{\Sigma T} \geq 1 \\ 0 & pAT_{ij}^{T+1} + \Delta AT_{ij}^{\Sigma T} \leq 0 \\ pAT_{ij}^{T+1} + \Delta AT_{ij}^{\Sigma T} & 0 < pAT_{ij}^{T+1} + \Delta AT_{ij}^{\Sigma T} < 1 \end{array} \right\} \tag{3}$$

where $pAT_{ij}^{T+1} : \mathbb{R} \rightarrow [0,1]$ is the most recent advertisement from agent j about service i, and $\Delta AT_{ij}^{\Sigma T}$ (AT-Discrepancy) is the discrepancy between advertisements and experiences obtained in the past (until time step T inclusive).

AT-Discrepancy $\Delta AT_{ij}^{\Sigma T}$: measures the discrepancy between the past advertisements made by agent j about service i and the experiences obtained when that service was requested.

$$\Delta AT_{ij}^{\Sigma T} = \frac{\sum_{t=0}^{T} \varphi(t,T)(pDT_{ij}^{t} - pAT_{ij}^{t})}{\sum_{t=0}^{T} \varphi(t,T)} \tag{4}$$

where $pAT_{ij}^{t} : \mathbb{R} \rightarrow [0,1]$ is the Partial Advertisements-based Trust for agent j, service i and time step t, and $\varphi(t,T)$ is a time forgetting function.

Note that $\Delta AT_{ij}^{\Sigma T} \in [-1,1]$, since $pDT_{ij}^{t}, pAT_{ij}^{t}, \varphi(t,T) \in [0,1]$ by definition Positive values of $\Delta AT_{ij}^{\Sigma T}$ means that the experiences obtained from agent j and service i were better than advertised, negative values have the opposite meaning, and zero means that the experiences matched perfectly with the advertisements.

AT Confidence(ATC_{ij}^{T+1}): assesses the degree of reliance of the Advertisements-based Trust as an estimation of the Quality of Service i to be obtained from agent j in the next time step.

$$ATC_{ij}^{T+1} = ITM_{ij}^{AT} \otimes (1 - \upsilon_{at}(\varDelta AT_{ij}^{t})) \tag{5}$$

where ITM_{ij}^{AT} is the intimacy level for AT, $\varDelta AT_{ij}^{t} = pDT_{ij}^{t} - pAT_{ij}^{t}$ is the partial discrepancy observed between AT and DT in time step t, $\upsilon_{at} \in [0,1]$ is a measure of the variability of $\varDelta AT$, and $\otimes$ is a T-norm operator.

As we have done for Direct Trust and Advertisements-based Trust, we define both partial and historic Recommendations-based Trust (RT). However, RT must handle the fact that there are potentially many providers of information (recommenders) about any other agent. As a result, we have to distinguish between the trust component due to the recommendations provided by a single agent and the trust component due to the recommendations provided by several agents; herein the latter is referred to as combined recommendation.

Recommendations-based Trust(RT_{ijk}^{T+1}): assesses the Quality of Service i expected from agent j in the next time step $(T+1)$, based on the recommendations from agent k.

$$RT_{ijk}^{T+1} = \left\{ \begin{array}{cc} 1 & pRT_{ijk}^{T+1} + \varDelta RT_{ijk}^{\Sigma T} \geq 1 \\ 0 & pRT_{ijk}^{T+1} + \varDelta RT_{ijk}^{\Sigma T} \leq 0 \\ pRT_{ijk}^{T+1} + \varDelta RT_{ijk}^{\Sigma T} & 0 < pRT_{ijk}^{T+1} + \varDelta RT_{ijk}^{T+1} < 1 \end{array} \right\} \tag{6}$$

where $pRT_{ijk}^{t} : \mathbb{R} \rightarrow [0,1]$ is the partial Recommendations-based Trust for agent j and service i obtained from agent k, and $\varDelta RT_{ijk}^{\Sigma T}$ (RT-Discrepancy) is the discrepancy between past recommendations and experiences about agent i and service j.

RT-Discrepancy($\varDelta RT_{ijk}^{\Sigma T}$): measures the discrepancy between the past recommendations by agent k about agent j and service i, and the experiences obtained using that service, until time step T inclusive.

$$\varDelta RT_{ijk}^{\Sigma T} = \frac{\sum_{t=0}^{T} \varphi(t,T)(pDT_{ij}^{t} - pRT_{ijk}^{t})}{\sum_{t=0}^{T} \varphi(t,T)} \tag{7}$$

where $pRT_{ij}^{t} : \mathbb{R} \rightarrow [0,1]$ is the partial Recommendations-based Trust for agent j, service i and time step t, and $\varphi(t,T)$ is a time forgetting function.

Note that $\varDelta RT_{ijk}^{\Sigma T} \in [-1,1]$, since $pDT_{ij}^{t}, pRT_{ij}^{t}, \varphi(t,T) \in [0,1]$ by definition Positive values of $\varDelta RT_{ij}^{\Sigma T}$ means that the experiences obtained from agent j and service i were better than recommended, negative values have the opposite meaning, and zero means that the experiences matched perfectly the recommendations.

Combined Recommendations-based Trust(cRT_{ij}^{T+1})**:** assesses the Quality of Service i expected from agent j in the next time step, based on both historic information and the most recent recommendations about that service.

$$cRT_{ij}^{T+1} = \frac{\sum_{k=1}^{N_k} (RT_{ijk}^{T+1} \times RTC_{ij}^{T+1})}{\sum_{k=1}^{N_k} RTC_{ijk}^{T+1}} \tag{8}$$

where RT_{ijk}^{T+1} is the Recommendations-based Trust about agent j and service i obtained from agent k's recommendations, and RTC_{ijk}^{T+1} is the confidence on that belief as an estimation of the Quality of Service i to be obtained from agent j in $T + 1$.

The Combined Recommendations-based Trust aggregates the recommendations obtained from several agents. Simmilarly, the confidence on cRT is defined as an aggregation of the confidences on every recommendation.

RT Confidence(RTC_{ij}^{T+1})**:** assesses the degree of reliance of the Recommendations-based Trust (RT_{ijk}^{T+1}) obtained from agent k, as an estimation of the Quality of Service i to be obtained from agent j in the next time step.

$$RTC_{ijk}^{T+1} = ITM_{ij}^{RT} \otimes (1 - \upsilon_{rt}(\Delta RT_{ijk}^{t})) \tag{9}$$

where ITM_{ij}^{RT} is the intimacy level for RT, $\Delta RT_{ijk}^{t} = pDT_{ij}^{t} - pRT_{ijk}^{t}$ is the partial discrepancy observed between DT and RT in time step t, $\upsilon_{rt} \in [0, 1]$ is a measure of the variability of ΔRT_{ijk}^{t}, and $\otimes$ is a T-norm operator.

Combined RT Confidence($cRTC_{ij}^{T+1}$)**:** assesses the degree of reliance of the Combined Recommendations-based Trust as an estimation of the Quality of Service i to be obtained from agent j in the next time step.

$$cRTC_{ijk}^{T+1} = \bigoplus^{k}(RTC_{ijk}^{T+1}) \tag{10}$$

where $\bigoplus^{k}$ denotes the aggregation of the confidence associated to each recommender ($RTC_{ijk}^{\Sigma T}$) using a *T-conorm* operator.

Up to now we have defined the components of trust according to the source of information. Now we provide a global measure of trust that integrates the three components into a single belief: the Global Trust.

Global Trust(GT_{ij}^{T+1})**:** assesses the Quality of Service i expected from agent j during the next time step, using all the sources of information.

$$GT_{ij}^{T+1} = \frac{DT_{ij}^{\Sigma T} \times DTC_{ij}^{\Sigma T} + AT_{ij}^{T+1} \times ATC_{ij}^{T+1} + cRT_{ij}^{T+1} \times cRTC_{ij}^{T+1}}{DTC_{ij}^{\Sigma T} + ATC_{ij}^{T+1} + cRTC_{ij}^{T+1}} \tag{11}$$

where $DT_{ij}^{\Sigma T}$ is the Direct Trust for service i and agent j; AT_{ij}^{T+1} is the Anticipatory Advertisements-based Trust; cRT_{ij}^{T+1} is the Combined Recommendations-based Trust, and $DTC_{ij}^{\Sigma T}$, ATC_{ij}^{T+1}, RTC_{ij}^{T+1} are the confidences associated to DT, AT and cRT respectively.

Global Trust Confidence(GTC_{ij}^{T+1}): assesses the reliance of the Global Trust GT_{ij} as an estimation of the Quality of Service i to be obtained in the next time step.

$$GTC_{ij}^{T+1} = DTC_{ij}^{\Sigma T} \oplus ATC_{ij}^{T+1} \oplus cRTC_{ij}^{T+1} \tag{12}$$

where $\oplus$ is a *T-conorm* operator.

Remark that Global Trust and Global Trust Confidence can be used either independently or combined into a single value (eg. $GT \times GTC$), depending on the specific application domain.

3 Experiments

We have chosen the ART Testbed [8] to test our test model. The ART Testbed is a simulator of the *art appraisals* domain whose goal is twofold: to serve as a competition forum in which researchers can compare their technologies against objective metrics, and as an experimental tool, with flexible parameters, allowing researchers to perform customizable, easily-repeatable experiments. In the art appraisal domain, agents act as painting appraisers with varying levels of expertise in different artistic eras (e.g. classical, impressionist, postmodern). Clients request appraisals for paintings from different eras. Appraisers can use both their own opinions and opinions purchased to other agents, so as to make more accurate appraisals. Appraisers estimate the accuracy of the opinions they send by the cost they choose to invest in generating an opinion, but they may lie about the estimated accuracy of their opinions. Appraisers receive more clients, and thus more profit, for producing more accurate appraisals. Appraisers may also purchase reputation information from other agents. The decisions about which opinion providers and reputation providers to trust strongly impact the accuracy of their final appraisals. In competition mode, the winning agent is selected as the appraiser with the highest bank account balance, which depends basically on the ability of an agent to (1) estimate the value of its paintings most accurately and (2) purchase more valuable information.

It is easy to map our trust model to the ART Testbed domain because it uses continuous variables and includes both advertisements (named certainties) and recommendations (named reputations). Purchased opinions about the value of a painting are the source of experience used to calculate DT, reputations are mapped to recommendations, and finally, advertisements are mapped to certainties, which are values provided by an agent about the accuracy of its opinions. Finally, the concept of a weight in the ART refers to a global measure attached to an agent to represent their opinion's accuracy. In our experiments, we use Global Trust $\times$ Global Trust Confidence to obtain those weights.

In order to evaluate the gains and drawbacks of using an anticipatory trust model in dynamic and uncertain environments, we have compared three models to handle trust and reputation: *anticipatory*, *non anticipatory without honesty*, and *non anticipatory with honesty.* The anticipatory model implements the trust model described in this paper, while the non anticipatory models use only historic information to calculate the global trust; the one with honesty uses the discrepancy between information and experience to calculate the confidence on trust, while the one without honesty simply ignores such discrepancy.

We compare the three trust models introduced above along four variables, namely: *number of appraisals*, *average error*, *bank balance*, and *stability.* The number of appraisals (NA) measures the total number of appraisals obtained during an entire simulation, the average error (AE) is the mean of those appraisal's error, the bank balance (BB) is the difference between the revenues and the expenses, and the stability (ST) is the number of time steps in which the average error for the last 5 time steps changes less than a given criterium ($|\text{average error increment}| < 0.01$).

Since in the ART Testbed agents can use their own opinions to appraise a painting, and they know themselves very well, self-opinions tend to neutralize the influence of the opinions purchased to other agents. In order to remark the differences between the three trust models being compared, we have enforced all agents to use solely the opinions purchased to other agents, and not their own opinions.

We have conducted three groups of experiments: (a) experiments with *dynamic prices* following a market-like evolving process, (b) experiments varying the degree of *deception* (dishonesty), and (c) experiments varying the prices randomly (a parameter called noise establishes the maximum price variation per time step). The same experimental situation is used as the baseline for the three groups of experiments: an static scenario where agents always provide the best opinions they are able to obtain and are completely sincere. Each experiment varying a parameter is repeated twice. A single experiment involves 9 agents competing during 60 time steps, with the same proportion of agents (3) using each trust model. Each time step, there are 270 paintings belonging to 10 artistic eras to be distributed among appraiser agents according to their relative average error in the previous time step.

Figure 1 summarizes the results of the experiments varying prices dynamics, according to a market-like model consisting of alternating inflation/deflation periods. There are three experimental situations, from left to right: static prices, *slow* price dynamics, and *fast* price dynamics. The bars with diagonal lines represent the average score for the agents using the anticipatory model (ant), dotted bars for the non anticipatory model with honesty (na), and horizontal lines for the non anticipatory model without honesty(nh). Results show that all the agents perform similarly in case of a completely static environment with fixed prices

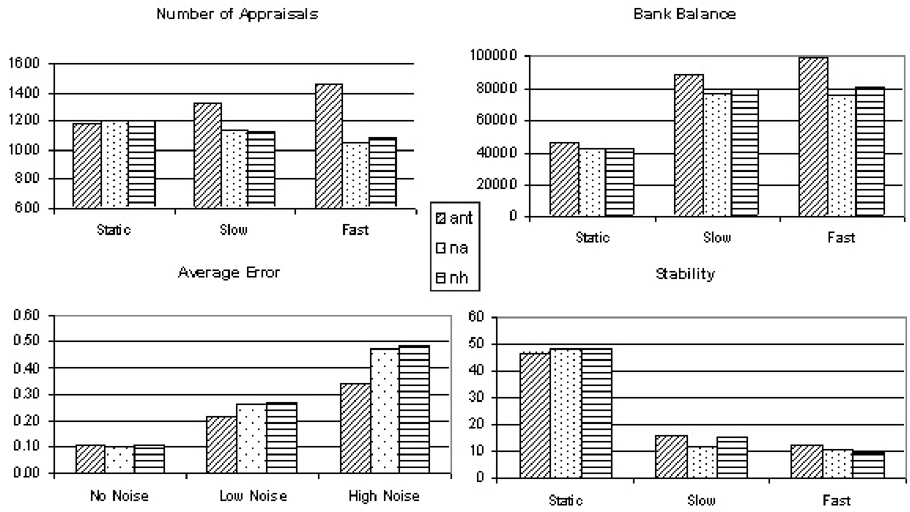

Fig. 1. Experiments with dynamic prices

and no deception. However, there are clear differences when considering dynamic prices, and these changes are stronger as prices change faster. In particular, the agents using the anticipatory trust model obtain the most accurate estimations of other agents(lower AE), achieve the most clients (higher NA), obtain the best economic results (higher BB) and remain stabilized the longest time, among the three models compared. Also to remark that the differences between the two non anticipatory models are very small as to be generalized.

Figure 2 sums up the results of our experiments varying the prices randomly according to certain degree of noise: *no noise* (static prices), *low noise*, and *high noise*. These results are completely consistent with the first group of experiments, the agents with the anticipatory model perform better in all the variables analyzed than the agents using the non anticipatory models.

Figure 3 shows the results of our experiments introducing certain degree of deception concerning both the advertisements (certainties) and the recommendations about the accuracy of agent opinions. We consider three experimental situations: *no deception*, *low deception*, and *high deception*. In this case, both the anticipatory and the non anticipatory models perform very similarly concerning the average error and the number of appraisals achieved, but the agents using anticipation achieve better economic results and remain stabilized for longer periods of time. Anticipatory agents earn more money even when obtaining fewer paintings to appraise, as has been the case for the low deception scenario. This result may seem contradictory because appraisals are the main source of revenues, but there is an explanation: the anticipatory trust model induced a more efficient behavior, in other words, anticipatory agents purchased fewer opinions to obtain appraisals of similar quality.

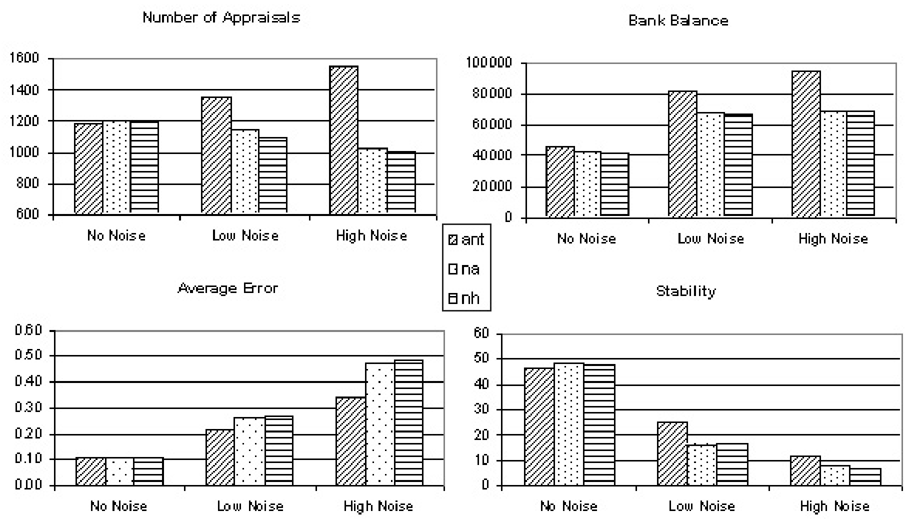

Fig. 2. Experiments with noise (random price changes)

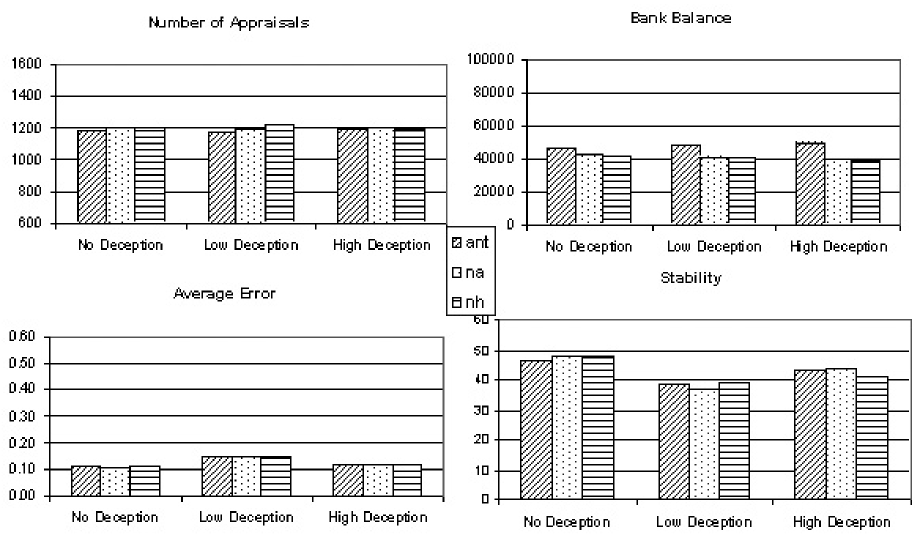

Fig. 3. Experiments with deception

4 Conclusions

In this paper we have introduced a computational model to handle trust in dynamic and uncertain domains such as electronic market places and distributed information systems. This model extends Rahman [1] notion of *semantic closeness* between experience and information to deal with continuous domains: first of all, we use continuous variables instead of discrete variables; second, our model combines the reputation information and the trust based on direct experiences (Rahman model uses only reputation); and third, we distinguish between two types of information: advertisements and recommendations.

Several frameworks to handle trust in agent societies rely upon the notion of honesty when considering the discrepancy between information and experience. Usually, the discrepancy observed between direct experience and information concerning that experience (witness information) is interpreted as a consequence of the information provider intentional behavior and is used to estimate the credibility (confidence) of that provider (more discrepancy implying less credibility); in other words, the discrepancy between information and experience is interpreted in terms of the honesty of the information provider.

Sabater [11] argues that although Rahmans approach is useful in some situations, it has some limitations because it is unable to differentiate between lying agents and agents that have a different view of the world. However, there are some reasons to adjust beliefs using the discrepancy between experience and information: on the one hand, in many domains, and specially in real applications, it is actually impossible to know whether an agent is lying or just thinking differently; on the other hand, it it is often more important to estimate the utility expected from an agent than figuring out whether an agent is lying or not.

In our model, the observed discrepancy between information and experience is used to adapt quicker to changes in the environment by anticipating changes in the world before experiencing them. This approach does not identify discrepancy with a bad behavior as such, instead our model uses that information to anticipate the future. However, in order to fully benefit from this approach, discrepancies between information and experience must be relatively consistent over time. That is to say, discrepancies between information and experience must follow a regular pattern so as to be useful.

We have used controlled experimental conditions to demonstrate the feasibility and utility of the anticipatory mechanism in a market-like simulation environment, the art appraisals domain. On the one hand, we have showed the utility of the anticipatory approach to adapt to changing environments, including both inflationary and deflationary dynamics. On the other hand, we have demonstrated the robustness of the model to deal with deception, both positive (over-valuating), to make an agent belief one is better than he actually is, and negative (under-valuating) deception, to make an agent belief a third agent is worst than itself.

Acknowledgements

Funded by projects CICYT TSI2005-07344, CICYT TEC2005-07 186 and CAM MADRINET S-0505/TIC/0255

References

1. S. H. A. Abdul-Rahman. Supporting trust in virtual communities. In 33^{th} *IEEE International Conference on Systems Sciences*, 2000.
2. S. Ba, A. Whinston, and H. Zhang. Building trust in online auction markets through an economic incentive mechanism. *Decision Support System*, 35:273–286, 2003.
3. S. Braynov and T. Sandholm. Trust revelation in multiagent interaction, 2002.
4. M. Butz, O. Sigaud, and P. Gérard, editors. *Anticipatory Behavior in Adaptive Learning Systems, Foundations, Theories, and Systems*, volume 2684 of *Lecture Notes in Computer Science*. Springer, 2003.
5. J. Carbo, J. Molina, and J. Davila. Trust management through fuzzy reputation. *International Journal of Cooperative Information Systems*, 12(1):135–155, Mar. 2003.
6. C. Castellfranchi and R. Falcone. Principles of trust for multiagent systems: Cognitive anatomy, social importance and quantification. In *Third International Conference on Multi-Agent Systems*, pages 72–79, 1998.
7. R. Conte and M. Paolucci. *Reputation in Artificial Societies.* Kluwer Academic Publishers, 2002.
8. K. Fullam, T. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K. S. Barber, J. Rosenschein, L. Vercouter, and M. Voss. A specification of the agent reputation and trust (art) testbed: Experimentation and competition for trust in agent societies. In *The Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2005)*, pages 512–518, 2005.
9. E. Lorini and C. Castelfranchi. The role of epistemic actions in expectations. In *Second Workshop of Anticipatory Behavior in Adaptive Learning Systems 2004 (ABIALS 2004)*, 2004.
10. S. Marsh. Trust in distributed artificial intelligence. In Castelfranchi and Werner, editors, *Lecture Notes in Artificial Intelligence 830*, pages 94–112. Springer Verlag, 1994.
11. J. Sabater. *Trust and Reputation for agent societies.* Consejo Superior de Investigaciones Cientificas, Bellaterra, Spain, 2003.
12. J. Sabater and C. Sierra. Regret: a reputation model for gregarious societies. In *Fourth Workshop on Deception, Fraud and Trust in Agent Societies*, pages 61–69, Montreal, Canada, 2001.
13. S. Wasserman and J. Galaskiewicz. *Advances in Social Network Analysis.* Sage Publications, Thousand Oaks, U.S., 1994.
14. B. Yu and M. Singh. A social mechanism for reputation management in electronic communities. *Lecture Notes in Computer Science*, 1860:154–165, 2000.
15. G. Zacharia and P. Maes. Trust management through reputation mechanisms. *Applied Artificial Intelligence*, 14:881–907, 2000.