A Computational Trust Model for Multi-Agent Interactions based on Confidence and Reputation

Sarvapali D. Ramchurn,
Nicholas R. Jennings
IAM group, ECS
University of Southampton, UK.
{sdr01r,nrj}@ecs.soton.ac.uk

Carles Sierra,
Lluís Godo
IIIA, CSIC
Bellaterra, Barcelona, Spain.
{sierra,godo}@iiia.csic.es

Abstract

In open environments in which autonomous agents can break contracts, computational models of trust have an important role to play in determining who to interact with and how interactions unfold. To this end, we develop such a trust model, based on confidence and reputation, and show how it can be concretely applied, using fuzzy sets, to guide agents in evaluating past interactions and in establishing new contracts with one another.

1 Introduction

Agents generally interact by making commitments to (contracts with) one another to carry out particular tasks. However, in most realistic environments there is no guarantee that a contracted agent will actually enact its commitments (because it may defect to gain higher utility or because there is uncertainty about whether the task can actually be achieved). In such situations, computational models of trust (here defined as the positive expectation that an interaction partner will act benignly and cooperatively in situations where defecting would prove more profitable to itself [Dasgupta, 1998]) have an important role to play. First, to help determine the most reliable interaction partner (i.e. those in which the agent has the highest trust). Second, to influence the interaction process itself (e.g. an agent's negotiation stance may vary according to the opponent's trust level). Third, to define the set of issues that need to be settled in the contract (i.e. the higher the trust, the more that can be left implicit in the contract).

Generally speaking, interactions go through three main phases; (i) a negotiation dialogue during which the terms of the contract are agreed upon and agents assign an expected utility to the contract, (ii) an execution phase during which there are opportunities for the contracted agent to defect, and (iii) an outcome evaluation phase where the client agent assesses the outcome of the task and finally derives some utility. In the cases where an agent has an incentive to defect, the client agent can judge whether the contractor is trustworthy by assessing its performance, relative to the initially contracted agreement, given its *perception of the prevailing task and the context*. Thus, the trust value for a specific agent for a specific task needs to take into account the potential utility loss or risk (associated with the task in question) in a contract given information about the context in which the contract is enacted [Marsh, 1994]. This follows from the fact that cooperating under high potential losses shows greater trustworthiness than otherwise [Yamagishi *et al.*, 1998].

Trust values, thus devised, can guide future contract negotiation in order to ensure that guarantees are provided against losses. Thus, if trust is sufficiently high, the contracted agent is deemed reliable. This means less time can be spent looking for potential contractors, negotiating about the minute guarantees present in the contract and, accordingly, giving more freedom to the contracted agent to enact its part of the deal. Conversely, when trust is low, the agents may spend a significant time specifying the guarantees associated with a contract or, if possible, avoiding future interactions with such agents.

Given this background, a number of computational models of trust have been developed (mainly based on theories from sociology). In [Marsh, 1994] for example, trust is taken to be a value between -1 and 1 that is calculated by taking into account risk in the interaction and the competence level of an interaction partner. However, these concepts are not given any precise grounding and they do not take into account past experience and reputation values of the contracted agent. In [Sabater and Sierra, 2002], reputation symbolises trust and competence levels are gathered from the social network in which the agents are embedded. The main value of this model lies in showing how reputation can be used to guide an agent's negotiation stance, but the evaluation of direct interactions is overly simple (disregarding utility loss and context). [Mui *et al.*, 2002] adopt a probabilistic approach to modelling trust that takes into account past encounters as well as reputation information. However, it is not obvious how the model can concretely guide an agent's decision making since the trust value is not associated to particular issues of the contracts that have been reneged upon. In a more realistic setting, Witowski et al. develop an objective trust measure from an evaluation of past performance [Witowski *et al.*, 2001]. However, their approach overly simplifies the trust modelling problem and avoids reputation measures which could have enhanced the performance of their agents.

In general, extant trust models fail to capture the individuality of an agent in assessing the reliability of an interaction partner. Most models also neglect the fact that agents interact according to the norms and conventions determined by the society or environment within which they are situated [Esteva *et al.*, 2001]. To this end, this paper develops a computational model of trust that rectifies these shortcomings.

By taking into account its past experience (from direct interactions) and information gathered from other agents (indirect interactions), an agent can build up beliefs about how trustworthy a contracted agent is likely to be in meeting the expected outcomes of particular contract issues (e.g. delivering goods on time or delivering high quality goods). In this respect, we conceive of two ways of assessing trustworthiness: (i) *Confidence* derived (mainly) from analysing the result of previous interactions with that agent, and (ii) *Reputation* acquired from the experiences of other agents in the community through gossip or by analysing signals sent by an agent. Both measure the same property; that is, the agent's believed reliability in doing what it says it will regarding particular issues of a contract. Here these measures are modelled using fuzzy sets to give agents a robust means of assessing the extent to which their interaction partners satisfy the issues of a contract. In particular, we advance the state of the art in the following ways. First, we delineate and computationally model context information, confidence measures, and risk in agent interactions. Second, we show how, using fuzzy sets, a measure of trust can be derived from the latter concepts and reputation information. Finally, we show how the trust measure can guide the choice of interaction parties, the stance that is taken during negotiation, and the issues that need to be agreed upon in a contract. The rest of the paper is organised as follows. Section 2 introduces the basic notions that we use throughout the paper, while section 3 shows how we devise a confidence measure. Section 4 details the trust model itself. Section 5 indicates how the model can be used and section 6 concludes.

2 Basic Notions

Let Ag be the society of agents noted as $\alpha, \beta, ... \in Ag$. A particular group of agents is noted as $G \subseteq Ag$. $\mathcal{G}$ denotes a partition $\{G_1, G_2, \ldots, G_l\}$ of the society of agents into non-empty groups. That is, for all $G_i, G_j \in \mathcal{G}, G_i \cap G_j = \emptyset$, $\bigcup_i G_i = Ag$. Therefore, any agent belongs to one and only one group. $\mathcal{T}$ denotes a totally ordered set of time points (sufficiently large to account for all agent interactions) noted as $t_0, t_1, \ldots$, such that $t_i > t_j$ if and only if $i > j$. Contracts are agreements about issues and the values these issues should have. Let $X = \{x, y, z, \ldots\}$ be the set of potential issues to include in a contract, and the domain of values taken by an issue x, be noted as D_x (for simplicity we assume that all D_x are a subset of real numbers $\mathbb{R}$). We will note that issue x takes the value $v \in D_x$ as $x = v$. Thus, a contract is a set of issue-value assignments noted as $O = \{x_1 = v_1, x_2 = v_2, ..., x_n = v_n\}$ where $x_i \in X$ and $v_i \in D_{x_i}$. We consider that agents invariably interact within an electronic institution [Esteva *et al.*, 2001] which specifies norms and conventions of interactions and dictates (some) issue-value assignments of contracts (see section 2.1). From now on, we will refer to the agent devising the contract as the *manager* and the contracted agent as the *contractor*.

2.1 Rules Dictating Expected Issue Assignments

We now consider how expectations about X are built based upon the agreed contract and the social setting. Here, the former provides a clear statement of what is expected with respect to each issue, while the latter may also give rise to expectations but these are not explicitly stated in the contract itself. For example, a buyer agent α might expect seller agent β to deliver goods nicely wrapped up in gift paper as opposed to in a carton box. This clause may not have been specified in the contract as it is a common belief in the client's group that goods must be nicely wrapped up.

Thus, at execution time, the contractor may not satisfy the manager's (contracted or not) expectations because (i) it is not able to meet the expectations, (ii) it is not willing to meet the expectations, or (iii) it is not aware of the unspecified expectations. In any case, the non-satisfaction of expectations *directly* impacts on the trust the client has in the contractor [Molm *et al.*, 2000] (unless a satisfactory reason is given for poor performance, but this is not considered here).

Here we consider three basic sources of unspecified expectations: (i) *General rules* that all agents in the society possess in common, (ii) *Social rules* that all agents within a particular group have in common, and (iii) *Institutional norms* that all agents interacting within a particular electronic institution have in common. This classification is necessary when we evaluate the performance of a contracted agent (see section 3). In more detail, rules allow an agent to infer expected issue-value assignments from a contract. Rules will be written in the following logical-like expressions:

$$\textbf{If } x_1 = v_1 \wedge x_2 = v_2 \wedge ... \wedge x_m = v_m \textbf{ Then } x = v$$

meaning that if $(x_i = v_i) \in O$ for all $i = 1, \ldots, m$, then issue x's value is expected to be equal to v. We assume that x is not appearing in the premise of the rule. We note by $Rules$ the set of all possible rules written using the above syntax[1] over the set X of issues and corresponding domains of values.

Group G's social rules, noted as $SocRules(G) \subseteq Rules$, are those shared by all agents in G. General rules, noted $GenRules \subseteq Rules$, are those shared by all agents. Institutional norms, noted as $InstNorms \subseteq Rules$, specify those rules that are enforced by the institution within which the (negotiating) agents interact. These norms are accepted by the agents involved in the negotiation process.

Given a contract O, we can devise the set of all of a manager's expectations (unspecified and specified), such that, given a manager $\alpha \in G$, we compute the set O_{exp} of expected issue-value assignments from O as the set of all conclusions of the rules of agent α, $Rules(\alpha) = GenRules \cup SocRules(G) \cup InstNorms$, that have their premise satisfied by the equalities in the contract O. The complete expanded contract is therefore defined as $O^+ = O \cup O_{exp}$. For each issue therein, the manager will have a confidence that the expected values will actually be met.

2.2 Confidence

In measuring confidence in an issue, different managers may have different opinions of the reliability of a particular con-

[1] Richer syntaxes could also be thought of for premises in these rules, allowing for predicates like $\geq, \leq, \neq$.

tractor. We initially consider trustworthiness on a per issue basis given that agents may be more reliable in satisfying some issues than others. These measures of satisfaction on contractors' behaviours are not strictly probabilistic in nature (since they cannot be analysed just in terms of frequency of occurrence), but may depend on the individual view of the agent as well. We therefore choose a fuzzy set based approach to relate confidence levels with expected values for issues.

In our approach we assume that agents share a (small) set $\mathcal{L} = \{L_1, L_2, \ldots, L_k\}$ of linguistic labels to qualify the performance of a contractor on each issue. For instance, $\mathcal{L} = \{Bad, Average, Good\}$. These labels model the agent's view on the *possible* (approximate) deviations from the contractually signed values. For example, "good" for agent α may mean that contractor β delivers articles on time or that it will actually give a discount on the initially quoted price, but for agent γ, it means that β will deliver articles before the quoted time of delivery or that β will not overprice its article.

A manager α assigns a confidence level to each label L when modeling the performance of a contractor β over a particular issue x, noted as $C(\beta, x, L) \in [0,1]$ (we will not include the agent identifiers to simplify notation when the context permits). Intuitively, this models the manager's belief that the deviation of the contractor on that issue will be within the possible values determined by that label. For instance, a manager may express his belief that a contractor is "good" to a confidence level 0.8 in fulfilling the contractual values on price, "average" to a level of 0.4, and "bad" to a level of 0.

For each issue x, a manager defines the meaning of each label L by a fuzzy set over $\mathbb{R}$ as if 0 were the contractual value for the issue, and whose membership function is noted as $\mu_L^x : \mathbb{R} \to [0,1]$. And now, given an issue-value assignment in a contract, $(x = v) \in O$, the actual meaning of the label L with respect to x in the contract is represented by a new fuzzy set noted as L^* which is the shift of L by v, and whose membership function is $\mu_{L^*}^x(u) = \mu_L^x(u - v)$, where $u \in \mathbb{R}$. From this, we can compute the set of *expected* values determined by a confidence level $C(x, L)$ on a label L for issue x in contract O, noted as $EV_c(O, x, L)$, as those values whose membership degree to L^* is above the confidence level. That is,

$$EV_c(O, x, L) = \{u \mid \mu_{L^*}^x(u) \geq C(x, L)\} .$$

This is graphically[2] shown in figure 1. To illustrate the above concept, consider the following example; a manager is confident to the degree of 0.9 that the contractor delivers goods "on time" if the goods arrive 1 day late (or 1 day earlier). However, if the confidence in being "on time" is 0.5 (i.e. the uncertainty is larger), it might expect the goods to arrive 3 days late (or 3 days earlier).

Finally, having devised confidence values for each label of an issue, a manager can compute the set of expected values for that issue as the intersection of the expected values for each label. That is,

$$EV_c(O, x) = \bigcap_{L \in \mathcal{L}} EV_c(O, x, L) \quad (1)$$

[2]The shape of the membership function given only serves as an example. Arbitrarily complex functions can be used in reality.

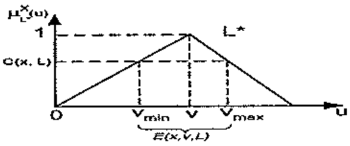

Figure 1: Confidence in a label L of an issue x.

In this context, an assignment of confidence values to labels in $\mathcal{L}$ is said to be consistent if $EV_c(O, x) \neq \emptyset$. This means that the manager always has a non empty set of expected values for an issue given its confidence values in different labels.

2.3 Reputation

Several models of reputation have been developed to illustrate how the transmission of (confidence) measures of reliability can be done [Yu and Singh, 2000; Sabater and Sierra, 2002]. Therefore, here we do not consider how this reputation information is gathered (and aggregated) from the other agents in the society. Rather, we assume this information is simply available from a social network that structures the knowledge that each agent has of its neighbours and keeps track of past interactions (as per [Sabater and Sierra, 2002]). Hence, we assume that a function $Rep : Ag \times X \times \mathcal{L} \to [0,1]$ exists where $Rep(\beta, x, L)$ represents the reputation degree of agent β with respect to the qualifying label L over issue x. (The name of the contractor will be omitted when the context unambiguously determines it.) The meaning of reputation here is an aggregation of opinions (confidence values in the previous sense) of some or all agents in Ag about one of them over a particular issue. Therefore, we can use reputation values in a similar way as before to compute the expected values for an issue x in a contract O, and label L as $EV_r(O, x, L) = \{u \mid \mu_{L^*}^x(u) \geq Rep(x, L)\}$ And, also similarly:

$$EV_r(O, x) = \bigcap_{L \in \mathcal{L}} EV_r(O, x, L) . \quad (2)$$

We initially assume that the assignment of reputation values for all labels are consistent such that $EV_r(O, x, L) \neq \emptyset$. This means that the manager always has a non-empty set of expected values for an issue given its reputation values in different labels.

2.4 Combined Range of Expected Values

Using just confidence levels or just reputation values to compute the set of expected values for a given issue x can be useful in certain extreme contexts (e.g. when the contractor does not rely at all on societal information, or when it has no other choice than to fully rely on it). However, in most situations, we consider that both sources of information should be taken into account in order to come up with a sensible set of expected values, applicable not only in the above two contexts but also to intermediate situations. In these situations we may want to consider a combination of both measures $CR : Ag \times X \times \mathcal{L} \to [0,1]$, which is, in the simplest case, a weighted average of both kinds of degrees (as in the previous cases we omit references to the agent whenever possible):

$$CR(x, L) = \kappa \cdot C(x, L) + (1 - \kappa) \cdot Rep(x, L), \quad (3)$$

where κ and $1-\kappa$ are suitable weightings that aim to model the relative importance of both information sources. This final range is computed as in the previous cases by first computing the expected values for an issue x and label L as $EV_{cr}(O,x,L) = \{u \mid \mu^{x}_{L^*}(u) \geq CR(x,L)\}$, and then the intersection of the expected ranges for all the labels $L \in \mathcal{L}$:

$$EV_{cr}(O,x) = \bigcap_{L\in\mathcal{L}} EV_{cr}(O,x,L)\,. \qquad (4)$$

We shall return to the issue of the weights in section 4, where we illustrate how reputation and confidence values ultimately lead us to devising actual trust values. However, the calculation of confidence (and indirectly trust) needs to take into account the context within which interactions take place since it is the context that determines the risk associated with each interaction.

3 Computing Confidence and Context

3.1 Interaction Context

Context, by definition, is the setting in which an interaction between a manager and a contractor takes place. Generally speaking, a setting can cover anything that relates to the interaction, but here we restrict[3] it to the rules that apply to the contract being negotiated (section 2.1) together with the agent's previous interaction experience.

In more detail, institutional norms and general rules are shared rules that form the objective context, while the subjective context is formed by the social rules of the group to which the manager α belongs and its interaction history with a given contractor β. Rules were defined in section 2.1 and history can be understood as the set of precedent cases. Each manager agent is assumed to have a utility function (applying over the domain of values) defined as $U_x : D_x \rightarrow [0,1]$ for each issue x. Utility values will be used as a common scale to compare and aggregate information from past cases.

Given that each interaction takes place through a contract, each case records information about the initially agreed upon issue-value pairs of a contract O, the actual results after execution, O', the confidence levels of α on the contractor agent β, $\{C(\beta,x,L)\}_{L\in\mathcal{L}}$, the reputation degrees of β, $\{Rep(\beta,x,L)\}_{L\in\mathcal{L}}$, and the time, t, at which the contract was executed. Thus each case is represented as a tuple:

$$case = \langle O, O', \{C(\beta,x,L)\}_{L\in\mathcal{L}}, \{Rep(\beta,x,L)\}_{L\in\mathcal{L}}, t\rangle$$

and the history of interactions as a case base $CB = \{case_1, case_2, \ldots\}$.

Given the case base, the set of utility functions for each issue in the contract, and the set of rules that apply to the interaction between the agents α and β, α's context within which a new contract is executed is represented as the set:

$$\Sigma_{\alpha,\beta} = \langle CB, \{U_x\}_{x\in X}, Rules(\alpha)\rangle.$$

This context is dynamic since rules or cases can be added or removed over time (and/or utilities change).

[3]We believe these are necessary rather than sufficient features and future work will investigate exactly what other characteristics could usefully be incorporated.

3.2 Utility loss and confidence

From the context we can now infer the probability of a contractor defecting (hence causing some utility loss) from a proposed contract given the rules that apply to it. If, for example, most interaction partners are known to have defected many times in the previous contracts relating to the same issues being negotiated in a new contract, we can reasonably assume that there is a high probability that the agent chosen will defect again in future interactions. Conversely, if the interactions have been successful or more profitable than expected, there is low probability of making a loss.

Therefore, given a context $\Sigma_{\alpha,\beta}$ and a current agreed contract $\overline{O}$, for each issue x in $\overline{O}$, we can estimate, from the history of past interactions, a probabilistic distribution P of α's utility variation $\Delta U_x \in [-1,1]$ (negative or positive) relative to issue x. Values of ΔU_x correspond to the possible differences between the utility $U_x(v)$ of the agreed value $(x = v) \in \overline{O}$ and the utility $U_x(v')$ of the (unknown) final value $(x = v')$ in the executed contract $\overline{O}'$. Then we can say that the manager agent α has a certain *risk* with issue x when it estimates that $P(\Delta U_x < 0) > 0$. Of course, the higher this probability is, the higher the risk is (i.e. the higher the expected utility loss).

Therefore, we need to estimate the probability distribution[4] of ΔU_x, not only for those issues x appearing in $\overline{O}$, but also on the expanded contract $\overline{O}^+ = \overline{O} \cup \overline{O}_{exp}$ resulting from the application of the rules in the current context (see section 2.1). We have to do so analogously with the contracts in the precedent cases of the case base CB of the context. However, if we assume that the agreed contract is signed such that the norms of the institution $InstNorms$ under which the agents are operating are fully enforced (i.e. punishments are given for not acting by the norms), then the risk is zero[5] for those issue-value assignments insured by institutional norms, even though the inference from previous interactions could suggest that the agent would defect. We then remove all these insured issues from the analysis.

Now, assume we have a probability distribution P for ΔU_x, and let $x = v$ be the signed issue value according to the contract $\overline{O}^+$. This allows us to compute confidence levels $C(x,L)$ for each $L \in \mathcal{L}$. In order to determine confidence levels $C(x,L)$ we initially need to determine a significantly representative interval $[\delta_1,\delta_2]$ for ΔU_x (e.g. such that $P(\Delta U_x \in [\delta_1,\delta_2]) = 0.95$). The latter denotes the change of utility interval which determines (under reasonable conditions) an interval of issue values around the agreed value v for x, $[v^-,v^+] = \{U_x^{-1}(\delta+u) \mid \delta \in [\delta_1,\delta_2], 0 \leq \delta+u \leq 1\}$, where $u = U_x(v)$. The interval $[v^-,v^+]$ is actually the range where agent α can expect (with high probability) to find the final value for issue x.

Finally, to calculate confidence levels $C(x,L)$ for each la-

[4]Several methods can be used to estimate these probability distributions, but the way the probability model is derived is not central to the trust model we wish to devise.

[5]This assumes that the institution fully insures against any losses. This assumption could be removed and a risk level determined according to the institutional rules as well.

bel $L \in \mathcal{L}$, we want the interval $[v^-, v^+]$ to coincide with the set of expected values $EV(O, x)$ as computed in section 2.2. Therefore the solution amounts to devising an inverse procedure to that of section 2.2 to come up with appropriate values $C(x, L)$ for each $L \in \mathcal{L}$ such that

$$[v^-, v^+] = \bigcap_{L \in \mathcal{L}} \{u \in D_x \mid \mu^x_{L^*}(u) \geq C(x, L)\} \quad (5)$$

The confidence values are updated whenever the ranges they define do not cover the range determined through the procedure above. The following example illustrates the above method. Assume an agent α has $P(\Delta U_x \in [-0.3, +0.1]) = 0.95$ and that a value $v = 100$ has been proposed for an issue x. Also, α has $U_x(v) = 0.4$. Therefore, it can expect $v^- = U_x^{-1}(0.4 - 0.3)$ and $v^+ = U_x^{-1}(0.4 + 0.1)$. Suppose $v^- = 75$ and $v^+ = 105$. This range determines the confidence level in the issue by having the confidence in each label cover as much of the range $[75, 105]$ as possible.

4 The Trust Model

As argued in section 1, there are two sources of information that permit an agent to build trust: confidence and reputation. We can therefore imagine three ways of defining trust (we consider the third one as the most appropriate):

1. **Trust = Confidence**

   In this case only the direct interaction is considered as a valuable source of information about the performance of another agent. The manager's first contract will be subject to total uncertainty and only after the number of interactions is significant will this way of defining trust start to work effectively. In cases where the trust in most agents of the society is low, this measure may be valid since the agent may not believe the opinions gathered to build a reputation value (e.g. due to known collusion among contractors and other managers).

   Given an issue-value assignment, $x = v$, present in a proposed contract, agent α's confidence levels $\{C(\beta, x, L)\}_{L \in \mathcal{L}}$ can be used (as shown in section 2.2) to determine the interval $EV_c(O, x)$ of expected values at the time of the contract execution. Therefore, the overall trust on issue x is defined through an inverse relation to the maximum utility loss when the value of issue x is varied on the interval $EV_c(O, x)$. In effect, if we let $\Delta^c_{loss} = \sup\{U_x(v) - U_x(w) \mid w \in EV_c(O, x)\}$ (sup is the least upper bound of the set), then we simply define the trust of the manager α in the contractor β over issue x as:

$$T^c(\beta, x) = \min(1, 1 - \Delta^c_{loss}) \quad (6)$$

2. **Trust = Reputation**

   Using confidence on its own, the trust in an agent becomes a useful measure only when there are a sufficiently large number of interactions. However, when the encounters have been scarce at the time of signing a contract, reputation information may be more useful [Mui *et al.*, 2002; Sabater and Sierra, 2002]. This is the usual case; when an agent first joins a society and decides to interact with its members, the only meaningful information it can have about the others' behaviour is their reputation, since it has not signed any contracts yet. Similar to the previous case, where trust = confidence, we can use the range $EV_r(O, x)$ to compute $\Delta^r_{loss} = \sup\{U_x(v) - U_x(w) \mid w \in EV_r(O, x)\}$ in order to define a trust value as:

$$T^r(\beta, x) = \min(1, 1 - \Delta^r_{loss}) \quad (7)$$

3. **Trust = Confidence and Reputation**

   As already mentioned in section 2.4, in most situations we believe it is preferable to consider both confidence and reputation. The rationale is that as the agent interacts with another agent more (and more), it will put correspondingly greater reliance on its personal confidence measures rather than reputation values (since personal interactions are deemed to be more accurate than information gathered from other agents, which might be subject to noise). Therefore, in our trust model we use the combined degrees $\{CR(x, v, L)\}_{L \in \mathcal{L}}$, as given by equation 3, to define the interval of expected values $EV_{cr}(O, x)$, that provides us with a maximum loss in utility $\Delta^{cr}_{loss} = \sup\{U_x(v) - U_x(w) \mid w \in EV_{cr}(O, x)\}$. Then, similarly to the previous two cases, we can define the overall trust on agent β over issue x as:

$$T(\beta, x) = \min(1, 1 - \Delta^{cr}_{loss}) \quad (8)$$

   The weight κ in equation 3 should reflect the number of past interactions between α and β. Therefore, given a context $\Sigma_{\alpha,\beta} = \langle CB, \{U_x\}_{x \in X}, Rules(\alpha) \rangle$, here we propose to define κ as $\kappa = \max(1, |CB|/\theta_{min})$, where $|CB|$ is the number of interactions in the context of both agents and θ_{min} is the minimum number of interactions above which only the direct interaction is taken into account (other types of functions could be used as well).

Our three definitions (especially the last one) take trust to be a dynamic and rational concept relating past experience and reputation to newly contracted values (i.e. depending on the situation [Marsh, 1994; Molm *et al.*, 2000]).

Depending on the environment, a specific definition of trust will be chosen. In any case, we can now define the trust $T(\beta, X')$ of the manager agent α on a contractor agent β over a particular set $X' = \{x_1, ..., x_k\}$ of issues appearing in the contract $\overline{O}$ (or in the expanded one $\overline{O}^+$) as an aggregation of the trust in each individual issue (e.g. trust in delivering on time, paying on time and the product having the quality specified in the contract). That is, we postulate

$$T(\beta, X') = g(T(\beta, x_1), ..., T(\beta, x_k)) \quad (9)$$

where $g : [0, 1]^k \rightarrow [0, 1]$ is a *suitable* aggregation function[6]. If some issues are considered to be more important than others, the aggregation function should take this into consideration, for instance by means of different weights[7] given for

[6] Generally, an aggregation function is defined as a monotonic function such that $\min(u_1, ..., u_k) \leq g(u_1, ..., u_k) \leq \max(u_1, ..., u_k)$ (see [Calvo *et al.*, 2002] for a survey).

[7] Most aggregation operators are defined parametrically with respect to weights assigned to each component to be aggregated.

each issue $x_i \in X'$ where some are considered to be more important than others. A typical choice would be to take the aggregation[8] function as a weighted mean:

$$T(\beta, X') = \sum_{x_i \in X'} w_i \cdot T(\beta, x_i) \qquad (10)$$

where $\sum w_i = 1$ and $0 \leq w_i \leq 1$.

5 Using the Trust Model

The trust model we have built is not meant to stand by itself. Therefore, in this section we illustrate how our model can be effectively used to guide negotiations and aid agents in forming profitable groups.

5.1 Choosing Interaction Partners

When an agent, say α, has a particular task to contract for, it will decide on the issues to be negotiated and identify possible interaction partners, say $\{\beta_1, \beta_2, ..., \beta_p\} \subseteq Ag$. For each agent in this set, we can calculate the trust value for each issue (as per equations 6, 7, or 8) and aggregate those to give a general trust value for each agent (using equation 9). That is, $T(\beta_1, X'), T(\beta_2, X')...T(\beta_p, X')$, where $X' \subseteq X$ is the set of issues under consideration. Trust can thus provide an ordering of the agents in terms of their overall reliability for a proposed contract. The client agent can then easily choose the preferred agent or set of agents it would want to negotiate with (i.e. by choosing the most trustworthy one(s)).

5.2 Redefining Negotiation Intervals

At contracting time, issue-value assignments, $x = v$, are agreed upon. Agents accept values that lie within a range $[v_{min}, v_{max}]$, such that $U_x(v_{min}) > 0$ and $U_x(v_{max}) > 0$. This interval is the acceptable range which it uses to offer and counter offer (according to a strategy) during negotiation [Jennings *et al.*, 2001]. On the other hand, given a potential issue-value assignment $x = v$ in an offer, an agent can compute, as shown in section 2, an interval of expected values $EV(O, x) = [ev^-, ev^+]$ over which the value v' obtained after execution can vary (as shown in equations 1, 2 or 4). This range defines the uncertainty in the value of the issue and if the acceptable range $[ev^-, ev^+]$ does not fit within $[v_{min}, v_{max}]$, there exists the possibility that the final value lies outside the acceptable region.

Therefore, the agent can restrict the negotiation interval $[v_{min}, v_{max}]$ with respect to the set of expected values $[ev^-, ev^+]$ as shown below. We first define the set of possible contracts, $\bar{O}_x$, that are consistent with the expected values of x and its acceptance range, and then define the corrected values for v_{min} and v_{max}:

$$\bar{O}_x = \{O \mid (x = v) \in O, EV(O, x) \subseteq [v_{min}, v_{max}]\}$$
$$v'_{min} = \inf\{v \mid (x = v) \in O, O \in \bar{O}_x\}$$
$$v'_{max} = \sup\{v \mid (x = v) \in O, O \in \bar{O}_x\}$$

This will shrink the range of negotiable values (i.e. from $[v_{min}, v_{max}]$ to $[v'_{min}, v'_{max}]$) to ensure that the final outcome fits within the range $[v_{min}, v_{max}]$ (depending on the utility obtained by the agent due to high or low values of the issue x, it can decide to stick to one of the two limits defined in $[v_{min}, v_{max}]$). On the one hand, this can help the agent reduce the time to negotiate over the value of each issue (e.g. if the range is small, the number of offers possible is also small), while on the other hand, the manager can make better decisions in parallel with the negotiation (e.g. if the contractor is expected to deliver 1 day late rather than in the agreed 3 days, the manager can instead agree on 4 days and adjust its other tasks to fit with the new delivery date). The above operation can also allow the agent to achieve better deals (e.g. if the redefined range specifies high utility values of the issue the negotiation strategy, can be adjusted in order to concede less than it would otherwise).

5.3 Extending the Set of Negotiable Issues

Initially we argued that higher trust should reduce the negotiation dialogue and lower trust should increase the number of issues negotiated over. In this section we deal with this particular use of trust in defining the issues that need to be negotiated over. To this end, issues not explicitly included in a contract may receive an expected value through one of the rules mentioned before:

$$\mathbf{r} : \textbf{If } x_1 = v_1 \wedge x_2 = v_2 \wedge ... \wedge x_m = v_m \textbf{ Then } x = v$$

Thus, if the premise of such a rule is true in a contract, the issue x is expected to have the value v. In case the trust in the agent fulfilling the premises is not very high, it would mean that the values $v_1, v_2, ..., v_n$ may not be eventually satisfied. In such a case, to ensure that the issue x actually receives value v it should be added to the negotiated terms of the contract. Formally, this means that if $T(\beta, X_r) < threshold$, ($T(\beta, X_r)$ defined as per equation 9), and where X_r is the set of issues in the premise of rule r, then the issue x in the conclusion of the rule should be added to the set of contract terms. On the other hand, as an agent becomes more confident that its interaction partner is actually performing well on the issues in the contract, it might eventually be pointless negotiating on the issue if the premises of the issue pre-suppose that the value expected will actually be obtained. This is, if $T(\beta, X_r) > threshold$, then the issue x in the conclusion of the rule can be removed from the set of contract terms.

The two processes described above serve to expand and shrink the space of negotiation issues. For a new entrant to the system, for example, the trust value others have in it would probably be very low and hence the number of issues negotiated over will be large. But, as it acquires the trust of others, the number of issues it would need to negotiate will go down. Ultimately, with more trust, the set of negotiable issues can thus be reduced to a minimal set, affording shorter negotiation dialogues. Conversely, with less trust, the negotiable issues expand, trading off the length of dialogues with better expected utility than otherwise.

6 Conclusions and Future Work

In this paper we have described the necessary components to build up a concrete computational trust model based on direct and indirect multi-agent interactions. We have instanti-

[8]More sophisticated aggregation models (based, for example, on different Lebesgue, Choquet, or Sugeno integrals) could also be used [Calvo *et al.*, 2002].

ated context, risk and confidence values using rules that apply over the issues negotiated in a contract. From these components a measure of trust has been proposed. Moreover, we have shown the worth of our model by describing how it can directly guide an agent's decisions in (i) choosing interaction partners, (ii) devising the set of negotiable issues, and (iii) determining negotiation intervals. The latter enable an agent to reduce time spent in negotiation dialogues, choose more reliable contractors and adapt the negotiation strategy to each contractor. These are not possible using current trust models. However, due to time limitations, it was not possible to simulate the use of the model in such a context and show the results.

Future work will focus on studying and refining the properties of the model for both cooperative and competitive settings through simulations. Also, we aim to enhance the context model to define more attributes of interactions (including network topology and task complexity). The trust measure will be made more sensitive to the stance taken by an opponent during the negotiation dialogue (e.g. if the opponent provided arguments backing its reliability). The variable κ will take into consideration such aspects. We will also consider in what respects the variable *threshold* can be calculated through the expected utility to be obtained from the values in those issues concerned. Finally we will investigate the properties of the model relative to the number of agents in the system and the number of interactions possible.

Acknowledgements

Sarvapali D. Ramchurn is supported by the FEEL project (IST-2000-26135) funded by the EU commission through the Disappearing Computer initiative. This work was also supported by the AgentCities mobility grant (IST-2000-28384).

References

[Calvo *et al.*, 2002] T. Calvo, A. Kolesarova, M. Komornikova, and R. Mesiar. Aggregation operators: Properties, classes and construction methods. In T. Calvo, G. Mayor, and R. Mesiar, editors, *Studies in Fuzziness and Soft Computing, Vol. 97*, pages 3–104. Physica-Verlag, Heilderberg, 2002.

[Dasgupta, 1998] P Dasgupta. *Trust as a Commodity, in Trust: Making and Breaking Cooperative Relations*, chapter 4, pages 49–72. Gambetta(ed), Blackwell, Department of Sociology, University of Oxford, 1998.

[Esteva *et al.*, 2001] M. Esteva, J. A. Rodríguez, C. Sierra, P. Garcia, and J. L. Arcos. On the formal specifications of electronic institutions. *Lecture Notes in Artificial Intelligence*, (1991):126–147, 2001.

[Jennings *et al.*, 2001] N. R. Jennings, P. Faratin, A. R. Lomuscio, S. Parsons, C. Sierra, and M. Wooldridge. Automated negotiation: prospects, methods and challenges. *Int. J. of Group Decision and Negotiation*, 10(2):199–215, 2001.

[Marsh, 1994] S. Marsh. *Formalising Trust as a Computational Concept*. PhD thesis, Department of Mathematics and Computer Science, University of Stirling, 1994.

[Molm *et al.*, 2000] L. D. Molm, N. Takahashi, and G. Peterson. Risk and trust in social exchange: An experimental test of a classical proposition. *American Journal of Sociology*, 105:1396–1427, 2000.

[Mui *et al.*, 2002] L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation. In *35th Hawaii International Conference on System Science (HICSS)*, 2002.

[Sabater and Sierra, 2002] J. Sabater and C. Sierra. Regret: a reputation model for gregarious societies. In C. Castelfranchi and L. Johnson, editors, *Proc. of the 1st Int. Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 475–482, 2002.

[Witowski *et al.*, 2001] M. Witowski, A. Artikis, and J. Pitt. Experiments in building experiential trust in a society of objective-trust based agents. In R. Falcone, M. P. Singh, and Y. Tan, editors, *Trust in Cyber-societies*, volume 2246 of *Lecture Notes in Computer Science*, pages 111–132. Springer, 2001.

[Yamagishi *et al.*, 1998] T. Yamagishi, K. Cook, and M. Watabe. Uncertainty, trust, and commitment formation in the united states and japan. *American Journal of Sociology*, pages 104:165–94, 1998.

[Yu and Singh, 2000] B. Yu and M.P. Singh. A social mechanism of reputation management in electronic communities. In *Cooperative Information Agents*, pages 154–165, 2000.