A Fuzzy Approach to Reasoning with Trust, Distrust and Insufficient Trust

Nathan Griffiths

Department of Computer Science, University of Warwick,
Coventry, CV4 7AL, UK
nathan@dcs.warwick.ac.uk

Abstract. Multi-agent systems are based upon cooperative interactions between agents, in which agents provide information, resources and services to others. Typically agents are autonomous and self-interested, meaning that they have control over their own actions, and that they seek to maximise their own goal achievement, rather than necessarily acting in a benevolent or socially-oriented manner. Consequently, interaction outcomes are uncertain since commitments can be broken and the actual services rendered may differ from expectations in terms of cost or quality. Cooperation is, therefore, an uncertain interaction, that has an inherent risk of failure or reduced performance. In this paper we show how agents can use trust to manage this risk. Our approach uses fuzzy logic to represent trust and allow agents to reason with uncertain and imprecise information regarding others' trustworthiness.

1 Introduction

Cooperation is the foundation of all multi-agent systems. Agents typically lack the knowledge, capabilities or resources needed to achieve their objectives alone, and it is through cooperation that they are able to function effectively. Individual agents provide information, resources and services to others in exchange for some form of payment. In order to achieve flexibility and robustness, agents are typically given the autonomy to control their own individual goals and actions. By definition, however, this autonomy implies that agents have control over how they cooperate. In particular, agents determine for themselves when to initiate cooperation or assist others, when to rescind cooperative commitments, and how to conduct cooperative tasks. Consequently, where agents cooperate any one of them may change the nature of their contribution, or even cease to cooperate, at any time. For example, an agent may choose to delay the provision of information, perform a processing task to a reduced quality, or simply fail to fulfil its commitments. Such failures are costly to the remaining cooperating agents since their goals may not be achieved, or not achieved as effectively (i.e. to a lower quality or with an increased cost).

On entering into cooperation an agent begins an uncertain interaction in which there is a risk of failure (or reduced performance) due to the decisions and actions of another. To function effectively, an agent needs to manage this risk.

M. Klusch, M. Rovatsos, and T. Payne (Eds.): CIA 2006, LNCS 4149, pp. 360–374, 2006.
© Springer-Verlag Berlin Heidelberg 2006

In this paper we show how agents can use the notion of trust, based on their individual experiences, to manage this risk by selecting appropriate interaction partners. Our approach uses fuzzy logic to represent trust and to allow agents to reason with uncertain and imprecise information. In addition to positive trust, we introduce the notions of negative trust (distrust) and insufficient trust (untrust and undistrust), and show how agents can use these concepts to increase the effectiveness of their interactions. Our approach to using fuzzy logic to represent trust was initially described in [4]. In this paper, we describe a refinement of the reasoning process, discuss how negative and insufficient trust can be incorporated into an agent's reasoning, and present our initial experimental results.

2 Background

2.1 Trust and Reputation

Trust and reputation are related, but distinct, concepts. The former represents an agent's *individual* assessment of the reliability, honesty etc. of another, while the latter is a *social* notion corresponding to a group assessment of such issues. Reputation is generally built by combining trust assessments (or recommendations) given by a group of agents to obtain a single value representing an estimate of reputation. The process of combining individual assessments or recommendations into a group notion generally requires agents to make their private assessments of others publicly available. Agents do not necessarily need to reveal the full details of their private assessments but they do need to reveal whether a given agent is considered trustworthy or not[1]. In some situations this can be undesirable from an individual's perspective, since it involves revealing private information that may reduce future effectiveness. For example, suppose that an agent α frequently cooperates with β, who reliably provides high quality and timely information. If α were to make its assessment of β's reliability and quality (i.e. its trustworthiness) public, then β may become overloaded and unreliable for α's future interactions.

In providing trust information to establish reputation, an agent might reduce the effectiveness of its own future interactions. For an agent to provide such information, there must be some intrinsic motivation for information sharing. In the absence of such a motivation, there will be insufficient information to assess reputation. There are also general issues with reputation concerning the subjectivity and context-specific nature of feedback [5]. Although in many situations the benefits of reputation might outweigh the individual cost of trust information sharing, it is useful in general to consider trust and reputation as separate, enabling agents to use trust without considering reputation.

Many of the existing applications of trust combine the notions of trust and reputation by using a global aggregation of individual trust into a reputation

[1] Other potential approaches, such as a polling mechanism, remove the need to make such private information public, however most existing approaches require the sharing of individual "trust recommendations".

assessment [13, 14, 15, 16]. Such research tends not to address how trust itself can be used in an individual's decision making. In this paper we focus specifically on trust from an individual's perspective, and do not consider reputation further. Our approach is complimentary to reputation-based models, and we view trust and reputation as both playing an important role in a complete system. Moreover, existing approaches also do not account for the roles of distrust and insufficient trust in the decision making process.

Existing models of trust can be categorised according to how trust is used: for achieving security or for enhancing quality of service [3]. In this paper we focus on the quality of service perspective to enable agents to maximise the "quality" of their interactions according to their current preferences.

2.2 Fuzzy Logic for Trust

Trust represents an individual's assessment of the reliability, honesty etc. of another, and the level of trust ascribed to an agent is based on the individual's experiences with that agent: positive experiences lead to positive trust and negative ones to distrust. However, although based on the known outcomes of previous experiences there is inherent uncertainty regarding the level of trust ascribed to an agent. For example, there is no guarantee that a previously reliable agent will continue to be so. Fuzzy logic offers the ability to handle uncertainty and imprecision effectively [12], and is therefore ideally suited to reasoning about trust. Inference using fuzzy logic copes with imprecise inputs, such as assessments of quality, and allows inference rules to be specified using imprecise linguistic terms, such as "very high quality" or "slightly late". Existing approaches have successfully used fuzzy logic to represent trust in multi-agent systems [8, 11] and peer-to-peer systems [14]. However, these existing techniques use trust as a means of establishing reputation, rather than focusing on individual trust in its own right. In this paper we aim to show how agents can enhance their interactions by using trust based on their individual experiences. Moreover, the existing approaches do not adequately consider the notions of negative and insufficient trust. We describe, in this paper, a method that uses fuzzy logic to make assessments about various aspects of trust, and allows agents to make decisions based on trust, distrust and insufficient trust. Before presenting our approach, however, we introduce some basic fuzzy concepts.

2.3 Basic Fuzzy Concepts

In classical set theory the membership of an object in a set is clearly defined: it is either a member or it is not. For example, a person of age 10 might be a member of the set *young*, and not of the set *old*. Such sets are required to have well-defined boundaries. However, the *concept* of young does not have a clear boundary, and in some contexts age 30 might be considered to be young, and not in others. *Fuzzy sets* are based on the notion of a *membership function*, $\mu(x)$, which defines the degree to which a fuzzy variable x is a member of a set. Full membership is represented by 1, and no membership by 0. The membership function $\mu(x)$

maps x into the interval $[0, 1]$. For example, age 35 might have a membership of 0.8 in a fuzzy set $\widetilde{y}$, representing young ages, and a 0.1 membership in the set $\widetilde{o}$ representing old ages. We use a tilde accent, $\tilde{x}$, to indicate that a set x is a fuzzy set. The *universe of discourse* of a fuzzy set corresponds to the range of values that are considered, such as $[0, 130]$ for age.

Fuzzy sets are used to define *terms* with respect to a *variable*. For example, the sets $\widetilde{y}$ and $\widetilde{o}$ define the terms $\widetilde{young}$ and $\widetilde{old}$ respectively, on the variable *age*. Terms can be subjected to *modifiers* (also called *linguistic hedges*), such as very or $\mathsf{slightly}$, which serve to modify, or hedge, the membership function from its original definition. The former example *concentrates* the membership function, while the latter *dilates* it. A discussion of the mathematical definition of such modifiers is beyond the scope of this paper, but we adopt Zadeh's definitions which follow the intuitive linguistic meanings [17]. For example, if we define $\alpha = \int_Y \frac{\mu_\alpha(y)}{y}$ then $\mathsf{very}\ \alpha = \int_Y \frac{[\mu_\alpha(y)]^2}{y}$ and $\mathsf{slightly}\ \alpha = \int_Y \frac{[\mu_\alpha(y)]^{0.5}}{y}$ (for further details see [12, 17]).

Relations between variables can be defined using *fuzzy inference rules* of the form:

$$\begin{array}{l} \textbf{if } \mathit{input}_1 \text{ is } [\mathsf{modifier}_1]\ \widetilde{\mathit{term}}_1 \textbf{ and } \mathit{input}_2 \text{ is } [\mathsf{modifier}_2]\ \widetilde{\mathit{term}}_2 \\ \quad \textbf{then } \mathit{output} \text{ is } \mathsf{modifier}'\ \widetilde{\mathit{term}'} \end{array}$$

which define the relationship between antecedents (input_1 and input_2) and consequent (*output*), described by terms $\widetilde{\mathit{term}}_1$, $\widetilde{\mathit{term}}_2$ and $\widetilde{\mathit{term}'}$ and optional modifiers $\mathsf{modifier}_1$, $\mathsf{modifier}_2$ and $\mathsf{modifier}'$. For example, we might have rules such as the following.

$$\begin{array}{ll} \text{(R1)} & \textbf{if } \mathit{age} \text{ is } \widetilde{\mathit{young}} \textbf{ and } \mathit{income} \text{ is } \mathsf{very}\ \widetilde{\mathit{high}} \\ & \quad \textbf{then } \mathit{customerPotential} \text{ is } \widetilde{\mathit{high}} \\ \text{(R2)} & \textbf{if } \mathit{age} \text{ is } \widetilde{\mathit{old}} \textbf{ and } \mathit{income} \text{ is } \widetilde{\mathit{low}} \\ & \quad \textbf{then } \mathit{customerPotential} \text{ is } \widetilde{\mathit{medium}} \end{array}$$

Rules are applied in parallel, and the conclusion membership degrees are aggregated by superimposing the resultant membership curves (i.e. by taking the fuzzy union of the resulting fuzzy sets). We adopt a Mamdani min-max approach to inference, such that the membership degree of the rule conclusions is clipped at a level determined by the minimum of the maximum membership values of the intersections of the fuzzy value antecedent and input pairs [7]. This ensures that the degree of membership in the antecedents is reflected in the output. We give an example of how Mamdani min-max inference operates for fuzzy trust in Section 5.

A crisp value can be obtained from the result of inference by *defuzzifying* the aggregated consequents. There are many methods for defuzzification, but for simplicity we take the centre of the area bounded by the membership curve. (Further discussion of the concepts introduced in this section can be found in [12].)

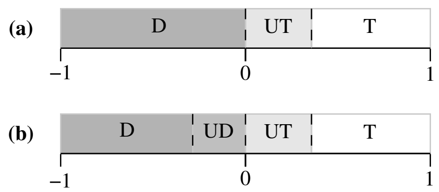

Fig. 1. (a) Marsh's notions of trust and, (b) our addition of undistrust, where D, UD, UT, and T correspond to distrust, undistrust, untrust, and trust respectively

3 Trust

Our proposed mechanism builds on existing work using service-oriented trust in agent-based systems. Trust is generally taken to be the belief that an agent will act in the best interests of another (i.e. that it will cooperate), even if given the opportunity to do otherwise (i.e. to defect) [1, 2]. When entering into cooperation an agent can use its trust of potential partners to evaluate the risk of failure. Most previous work on trust has concentrated the positive side of trust (analogous to assessing the extent to which an agent is reliable), and has largely ignored the notion of *distrust* (analogous to assessing the extent to which an agent is *un*reliable). Distrust is not simply the negation of trust [6], but rather it is an explicit belief that an agent will act against the best interests of another [9]. Alternatively, *untrust* corresponds to the space between distrust and trust, in which an agent is positively trusted, but not sufficiently to cooperate with. This view of trust, proposed by Marsh, is illustrated in Fig. 1(a). Marsh argues that distrust is an important concept, that can play an important role in an agent's reasoning, complimenting trust itself [9].

We concur with Marsh's view regarding the importance of distrust, and in this paper we provide a mechanism for agents to make use of distrust in their decision making. In addition to distrust, untrust and trust, however, we propose a new notion of *undistrust*. Untrust is defined as positive trust, but insufficient to support cooperation. For distrust to play a useful role in an agent's reasoning, we argue that a similar region of undistrust is needed, namely, a region of negative trust but insufficient to make definite conclusions in the reasoning process. Fig. 1(b) illustrates our definition of the notions of trust, distrust, untrust and undistrust. Although agents can not use untrust and undistrust to make definite conclusions regarding trust, they can still make use of the notions of untrust and undistrust in their reasoning regarding cooperation. For example, if there are no trusted agents with whom to interact then an agent may choose to interact with an untrusted, or even an undistrusted, agent provided that the cost of failure is relatively low (i.e. where it is better to have tried and failed that not to have tried at all).

4 Interaction Histories

Trust is based on an agent's individual experiences (since in this paper, as noted above, we do not consider reputation), and so each agent must keep track of its previous interactions. Some interactions may simply have a binary result of success or failure. However, typical cooperative interactions are more complex than simple succeed or fail tasks, and tasks may partially succeed, or be completed with different characteristics than expected. For example, an agent that has agreed to provide information may provide it late or with less detail than expected. This might not cause the goal of the receiving agent to completely fail, but it may cause the receiving agent's level of performance to reduce. Therefore, to enable agents to make effective use of trust we require them to represent more than a simple expectation about success or failure. We take a multi-dimensional view of trust as comprising the combination of the different dimensions of an interaction, such as the quality of a task or the cost imposed for executing it, in addition to whether an interaction was successful or not. Agents model such characteristics as *dimensions of trust*, which taken together give an assessment of an agent's trustworthiness. For illustrative purposes, in this paper we consider the dimensions of success, cost, and quality, although other dimensions are equally possible. The multi-dimensional approach provides a mechanism that allows agents to reason about the specific characteristics of an interaction where appropriate.

In order to assess trust an agent must evaluate its experiences in each of the trust dimensions. For each interaction, and in each dimension, an agent's expectations will have either been met or not met. Agents maintain a history of the interactions that they have had with each other agent, and track the number of successful and unsuccessful interactions for each dimension, in terms of whether their expectations were met. Thus, for each dimension, d, and agent that has been cooperated with, α, an agent maintains a value I_α^{d+} which corresponds to the number of interactions in which its expectations were met, and a value I_α^{d-} in which they were not met. From these values, the *experience*, e_α^d, in each dimension d, for each agent α, can be calculated as:

$$e_\alpha^d = \frac{I_\alpha^{d+} - I_\alpha^{d-}}{I_\alpha^{d+} + I_\alpha^{d-}}$$

Such experience values are the basic information from which an agent can assess the trustworthiness of others. They are crisp values in the interval $[-1, 1]$ and must be translated into fuzzy values in order to reason about trust. Experience values are based directly on an agent's interaction histories, and so they are not uncertain in themselves. Rather, the uncertainty for trust comes from a lack of information about other agents' future actions. Therefore, each experience value is fuzzified by translating it into a fuzzy value defined by the singleton fuzzy set whose membership function is 0 at all points except for e_α^d which has a membership of 1. Thus, the fuzzified experience is given by $E_\alpha^d = \mathit{fuzzySingleton}(e_\alpha^d)$.

4.1 Purging Old Interactions

Agents keep track of the outcomes of their interactions by using a window of experiences that is maintained for each other agent. This window is bounded, such that there is an upper limit on the number of interactions that are recorded for any agent. The interaction window acts as a first-in first-out queue, and when full it is the earliest experiences that are removed to be replaced by new ones. Over time, however, the information stored may become outdated if the environment (particularly in terms of the character of the other agents) has changed and previous experiences are no longer relevant. Agents may change, and an agent that was reliable previously may no longer be so. To address this problem an agent purges outdated experiences from its interaction windows after a certain predefined period. Thus, even if an interaction window is not full, the record of experiences will be removed over time.

The delay between the occurrence of an interaction and the removal of its record from the interaction window is called the *purge lag*, and has a direct influence on how quickly an agent's trust assessments respond to changes in its environment. A small purge lag means that interaction records do not persist for long and so the effect of previous experiences decays quickly and trust assessments respond quickly to changes. However, a small purge lag also reduces the extent of the experiences that can be used to determine trust. If the purge lag is too small there will be insufficient experiences on which to base trust, and any small perturbations in others' reliability and honesty will have a significant effect on trust. Conversely, a large purge lag avoids magnifying the effects of small perturbations in others' reliability, but increases the number of interactions that are required to react to changes in the environment. Thus, trust assessments are slow to respond to change.

In determining trust it is important that an agent has sufficient experience on which to calculate trust. We define the confidence level in the experience for a particular dimension as the total number of interactions on which it is based.

$$confidence_{\alpha}^{d} = I_{\alpha}^{d+} + I_{\alpha}^{d-}$$

If this confidence level is below a predefined threshold then either a value of untrust or undistrust will be ascribed (for the success dimension), or a default value will be used (for other dimensions) as described in the following section. Note that there may be different levels of confidence for different dimensions. For example, there are likely to be fewer interactions relevant to quality than success, since only successful interactions will contribute to the quality dimension whereas all interactions will contribute to the success dimension.

5 Fuzzy Trust

We define fuzzy terms for experience in each of the dimensions in which agents record their interactions, in our case success, cost, and quality. Fuzzy terms are defined in reference to fuzzy variables, and for experience we define fuzzy

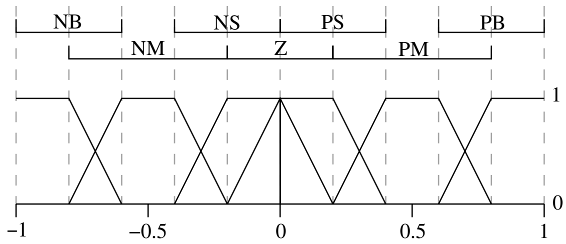

Fig. 2. Definition of fuzzy terms for experience, where NB, NM, NS, Z, PS, PM, and PB correspond to negative big, negative medium, negative small, zero, positive small, positive medium and positive big respectively

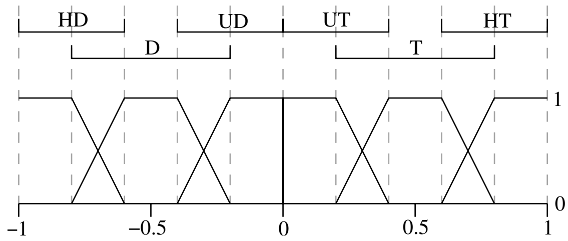

Fig. 3. Definition of fuzzy terms for trust, where HD, D, UD, UT, T, and HT correspond to high distrust, distrust, undistrust, untrust, trust, and high trust respectively

variables for each trust dimension. Thus, for our chosen dimensions we introduce E^s_α, E^c_α, and E^q_α corresponding to the experiences in the dimensions of success, cost, and quality for agent α respectively. The universe of discourse of these fuzzy variables is $[-1, 1]$, i.e. ranging from expectations never being met to expectations always being met. For each of these variables we define the terms: negative big, negative medium, negative small, zero, positive small, positive medium, and positive big. Other terms are possible, but these are sufficient for our purposes. The fuzzy sets that describe these terms are illustrated in Fig. 2.

In order to use fuzzy inference to determine trust, T_α, in an agent α we must also define trust as a fuzzy variable, with an associated set of fuzzy terms. The universe of discourse for trust is also $[-1, 1]$, i.e. complete distrust to complete trust, and we define the terms: high distrust, distrust, undistrust, untrust, trust, and high trust. These terms allow us to represent that there is insufficient trust for reasoning, in the form of untrust and undistrust, along with representing two degrees of trust and distrust. Again, other definitions are possible, but these are sufficient for our application. The fuzzy sets that describe these terms are illustrated in Fig. 3.

For each dimension we define a set of fuzzy inference rules that take the fuzzified experiences as antecedents and make conclusions regarding trust. The definition of these rules is the responsibility of the system developer, and we do not prescribe a particular rule set. In the experiments described in Section 6 we

($R_{UT}1$) **if** $\mathit{confidence}^d_\alpha < \mathit{minConfidence}$ **and** E^d_α is *positive* **then** T_α is $\widetilde{\mathit{untrust}}$
($R_{UT}2$) **if** $\mathit{confidence}^d_\alpha < \mathit{minConfidence}$ **and** E^d_α is *negative* **then** T_α is $\widetilde{\mathit{undistrust}}$
...
(R_T1) **if** E^d_α is $\widetilde{\mathit{negativeBig}}$ **then** T_α is $\widetilde{\mathit{highDistrust}}$
(R_T2) **if** E^d_α is $\widetilde{\mathit{negativeMedium}}$ **then** T_α is very $\widetilde{\mathit{distrust}}$ or $\widetilde{\mathit{undistrust}}$
(R_T3) **if** E^d_α is $\widetilde{\mathit{negativeSmall}}$ **then** T_α is $\widetilde{\mathit{undistrust}}$
(R_T4) **if** E^d_α is $\widetilde{\mathit{zero}}$ **then** T_α is $\widetilde{\mathit{undistrust}}$ or $\widetilde{\mathit{untrust}}$
(R_T5) **if** E^d_α is $\widetilde{\mathit{positiveSmall}}$ **then** T_α is $\widetilde{\mathit{untrust}}$
(R_T6) **if** E^d_α is $\widetilde{\mathit{positiveMedium}}$ **then** T_α is very $\widetilde{\mathit{trust}}$ or $\widetilde{\mathit{untrust}}$
(R_T7) **if** E^d_α is $\widetilde{\mathit{positiveBig}}$ **then** T_α is $\widetilde{\mathit{highTrust}}$
...
(R_Rn) **if** T_α is $\widetilde{\mathit{highTrust}}$ and F^c_α is $\widetilde{\mathit{medium}}$ and F^q_α is very $\widetilde{\mathit{high}}$ **then** R_α is $\widetilde{\mathit{high}}$
(R_Rm) **if** T_α is low $\widetilde{\mathit{distrust}}$ and F^c_α is $\widetilde{\mathit{medium}}$ and F^q_α is $\widetilde{\mathit{high}}$ **then** R_α is $\widetilde{\mathit{low}}$

Fig. 4. Example fuzzy inference rules

use the rules R_T1–R_T7 given in Fig. 4 along with additional rules of the form of R_Rn. Other rules are, of course, possible and can be easily incorporated into the system.

5.1 Determining Trust

Before determining the trustworthiness of an agent the assessor must check whether there have been sufficient previous interactions to calculate trust. All previous interactions will either have succeeded or failed (there is no notion of a 'partial' success), and so we use the success dimension to determine whether there is sufficient information to calculate trust. If there have not been sufficient interactions in the success dimension then the agent is ascribed a value of untrust or undistrust according to whether the interactions that have taken place are positive or negative overall, i.e. whether e^d_α is positive or negative. Thus, the first step in determining the trust of an agent α is to check whether there is sufficient confidence, i.e. that $\mathit{confidence}^s_\alpha \geq \mathit{minConfidence}$. If there is not sufficient confidence then trust T_α is defined by the fuzzy terms $\widetilde{\mathit{untrust}}$ or $\widetilde{\mathit{undistrust}}$ with a membership degree determined by the level of confidence and value of experience, as defined in rules $R_{UT}1$ and $R_{UT}2$. (Note that before firing rules $R_{UT}1$ and $R_{UT}2$ in fuzzy inference confidence is fuzzified in a similar manner to that described above for experience.)

Provided that there have been sufficient previous experiences, then fuzzy inference is used to calculate trust. To determine the trustworthiness of the potential interaction partners we must consider the inference rules for each of the trust dimensions. Each rule is considered in turn, and if there is a match between the input (i.e. E^d_α) and the fuzzy set defined by the antecedent of the rule, then the rule is fired. For example, if there is an overlap between the input E^d_α and the area defined by the term *negativeBig* then rule R_T1 is fired. If there is insufficient confidence in a particular dimension, the agent uses a default "experience" value

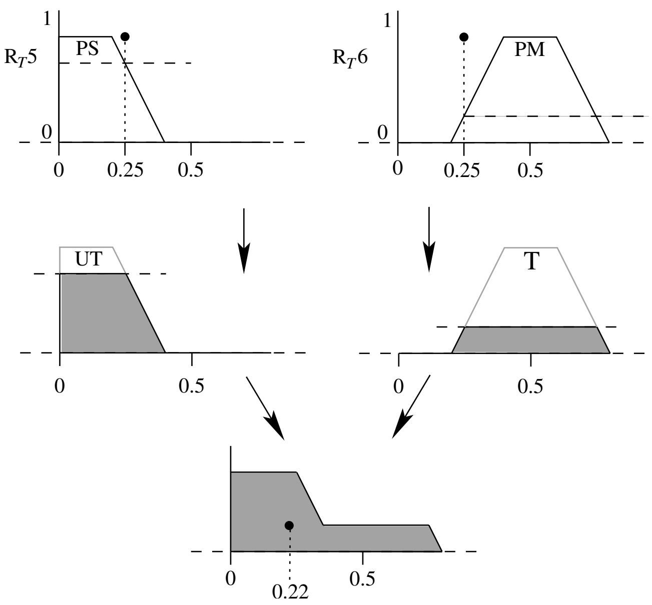

Fig. 5. A simple inference example showing the firing of rules R_T5 and R_T6

for that dimension, $default_d$. This value is determined by the the agent's trust disposition, with optimists using higher values than pessimists.

By way of example, suppose that for the success dimension we have determined that $E^s_\alpha = 0.25$ for agent α based on the experiences recorded in the interaction window. This crisp value is fuzzified as described above, and the fuzzy rules are then applied. In this case the input set matches with the antecedents of rules R_T5 and R_T6, i.e. *fuzzySingleton*(0.25) overlaps with the sets defined by the terms *positiveSmall* and *positiveMedium*. Using Mamdani min-max inference the membership of the conclusion fuzzy set is clipped by the degree of membership of the antecedent. The outputs of the rules are then aggregated by taking the fuzzy union. This is shown graphically in Fig. 5. The process is then continued for the other dimensions, with the outputs from any matching rules being combined with the existing output by taking the fuzzy union. Once rules R_T1–R_T7 have be applied for all dimensions we have determined a fuzzy value for trust T_α. A crisp value can be determined by defuzzifying as shown in Fig. 5, in this case resulting in a trust of 0.22.

5.2 Distrust

Once the potential cooperative partners have been ascribed trust values, the selecting agent can filter out all those that are distrusted. Since trust is a fuzzy value, checking for distrust is not a crisp operation, but instead involves considering the extent that trust is a member of the fuzzy set $\widetilde{highDistrust}$. A small membership in this set is (typically) insufficient to reject a partner whilst a high membership, indicating definite high distrust, should cause the agent to be rejected. Our approach is to check the *similarity*[2] of the fuzzy value $\widetilde{T_\alpha}$ with the fuzzy value whose membership function is defined solely by the $\widetilde{highDistrust}$ fuzzy set. This can be thought of as checking the similarity of the T_α and $\widetilde{highDistrust}$ membership graphs. If the similarity is above a threshold, *maxDistrust*, then the agent concerned is rejected, and is no longer considered to be a potential cooperative partner.

5.3 Untrust and Undistrust

If the trust level ascribed to an agent is untrust or undistrust, then the trust level is considered insufficient to directly make a decision regarding the agent's suitability (i.e. to reject the agent or to cooperate). Intuitively, each agent who is ascribed untrust or undistrust should not be directly considered for selection (although it should not be completely rejected either). However, in hostile or highly dynamic environments this can lead to problems, since all agents may be either distrusted and so explicitly rejected, or untrusted and undistrusted and so not considered for cooperation. This gives rise to deadlock. To avoid this situation we provide the facility for untrusted and undistrusted agents to be considered for a proportion of interactions. If there are no trusted agents that have the required capabilities then with some probability, called the *rebootstrap* rate, the agent with the highest trust level from the set of untrusted and undistrusted agents will be selected.

5.4 Selecting an Cooperative Partner

Assuming that there is a set of trusted (i.e. with a trust level above untrust) agents, then one of them can be selected for cooperation. Agents might simply use trust alone to select which agent to cooperate with by selecting the most trusted. However, typically there is additional information with which to make a decision. For example, each of the alternative agents may advertise a cost and quality for the interaction. In this case, the selecting agent can incorporate such information into its decision making. Since these advertised values represent uncertain information (i.e. the actual cost and quality are unknown at the point of making a decision), they also lend themselves to fuzzy inference. Thus, we introduce fuzzy rules that combine trust with each of the other decision factors and determine a rating for each alternative potential interaction partner. Each

---

[2] The experiments described in Section 6 are obtained using the NRC FuzzyJ Toolkit [10]. We adopt the definition of similarity given in FuzzyJ, namely:

$$\begin{aligned} \mathit{similarity}(a,b) = {} & \textbf{if } \mathit{necessity}(a,b) > 0.5 \\ & \textbf{then } \mathit{possibility}(a,b) \\ & \textbf{else } (\mathit{necessity}(a,b) + 0.5) \times \mathit{possibility}(a,b), \text{ where} \end{aligned}$$
$\mathit{necessity}(a,b) = 1 - \mathit{possibility}(\mathit{not}\ a, b)$, and
$\mathit{possibility}(a,b) = \mathit{max}_x(\mathit{min}(\mu_a(x), \mu_b(x)))$.

of these factors F_i is a crisp value, which can be fuzzified as a singleton set. We define a set of inference rules that have fuzzy trust and the fuzzy decision factors as antecedents and the *rating* for an agent as conclusions. These factors are domain specific. In our example, we use the advertised cost and quality from an agent α, denoted F_α^c and F_α^q respectively. Suppose that we have defined the fuzzy terms *low*, *medium* and *high* for these factors, according to the universe of discourse defined by the range of potential advertised cost and quality values. Similarly, suppose that we have terms *low*, *medium*, *high* and *reject* defined for ratings, which has a universe of discourse of $[0, 1]$. We then define a set of rules of form illustrated by R_Rn and R_Rm in Fig. 4. Rule R_Rn states that if an agent is trusted, has a medium advertised cost and a high advertised quality, then it has a high rating. Similarly, rule R_Rm states that if an agent is ascribed low distrust (insufficient distrust to cause a rejection), has a medium advertised cost and high advertised quality, then it has a low rating. Similarly to calculating trust, each of these rules is applied in parallel using Mamdani min-max inference, and a crisp rating value for agent α is obtained by defuzzifying the fuzzy rating. To balance the importance of the various decision factors (including trust), agents can scale the inputs before performing inference. For example, if cost is not currently important then the input E_α^c would be multiplied by some reduction factor, r, where $0 < r < 1$.

In order to select an agent to cooperate with, the selecting agent calculates the rating value for each alternative, and selects the one with the highest rating. After the interaction, the interaction window is updated according to whether the interaction was successful, and whether the expected (as determined by advertised value) cost and quality were met.

5.5 Bootstrapping

Initially agents have insufficient experience for reasoning. Therefore, each agent goes through a bootstrapping phase in which partners are chosen randomly by way of exploration. During this bootstrapping phase agents that are distrusted, undistrusted, untrusted, and trusted have an equal chance of being selected.

6 Experimental Results

Our approach has been validated experimentally, using the NRC FuzzyJ Toolkit [10] to implement the fuzzy decision mechanism. We constructed a test stub to generate the complete set of possible interactions that an agent might have, i.e. the outcome that would result for each choice of potential cooperative partner. Using this set we can then evaluate the effectiveness of different configurations of the decision mechanism for each set of possible interactions. Thus, we can make direct comparisons about how effective a given configuration of the fuzzy decision mechanism is given exactly the same set of possible interactions. In this section we describe initial results obtained by simulating an agent using fuzzy trust to select its cooperative partners in an environment of 50 others from whom

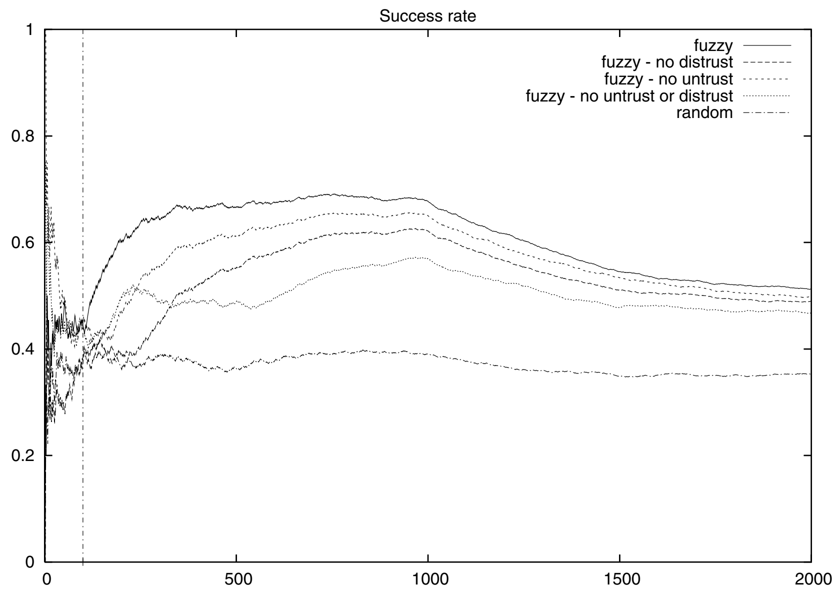

Fig. 6. The effect of reasoning with untrust and distrust on the success rate

a cooperative partner must be chosen each iteration (or cooperation avoided due to distrust or insufficient trust). The vertical line after 100 interactions signifies the end of the bootstrapping phase, and so to the left of this line agents are chosen randomly (from the set of agents that have the required capabilities). To the right of the bootstrapping line fuzzy trust is used to select a partner from the agents that have the required capabilities.

Fig. 6 shows the effectiveness of our decision mechanism on the success rate of cooperative interactions, and illustrates the usefulness of untrust and distrust. A random selection of cooperative partners is also shown for control purposes. The agents in the system were initially generated to be of average reliability, but were made less reliable after 1000 iterations. It can be seen that each of the fuzzy approaches give a significant increase in success rate. (Note that exactly the same generated "environment history" is used to obtain results for each approach.) Fig. 6 also shows that making explicit use of the notions of distrust, undistrust and untrust in decision making results in an increased success rate. The use of distrust alone (i.e. not using untrust and undistrust) gives a better result than using untrust and undistrust alone, and using positive trust alone (i.e. not using distrust, undistrust or untrust) gives the lowest success rate, although still significantly higher than a random selection.

The results shown in Fig. 6 are for a single generated "environment history". However, the actual success rate changes with the environment, since agents' reliability is different across environments. Fig. 7 shows the success rate and the rate that cost and quality expectations are met across a set of environments. The figures shown are averaged for 20 separate environments. It can be seen

Selection mechanism	Success rate	Cost rate	Quality rate
Fuzzy trust	0.64	0.73	0.84
Fuzzy trust (no distrust)	0.63	0.68	0.81
Fuzzy trust (no untrust or undistrust)	0.63	0.58	0.79
Fuzzy trust (no distrust, untrust or undistrust)	0.61	0.56	0.78
Random (control)	0.42	0.39	0.64

Fig. 7. Success rate and the rate that cost and quality expectations are met, averaged over a set of environments

that, as in Fig. 6, using trust gives a significant improvement in success rate over a random selection. Similarly, the use of distrust, undistrust and untrust also improves the success rate. When averaged over a set of environments the effect on the success rate of distrust, undistrust and untrust are similar. The use of fuzzy trust for decision making also has a significant effect on the rate at which cost expectations are met, with nearly a 35% improvement over a random selection. In the cost dimension, the use of untrust and undistrust has a greater effect than distrust, but the best results are again obtained by using distrust, untrust and undistrust together, giving a rise of 17% over using fuzzy trust that considers positive trust only (i.e. not using distrust, undistrust or untrust). We have obtained similar results in the quality dimension, as also shown in Fig. 7.

7 Conclusions

In this paper we have shown how fuzzy logic can be used to represent trust, and select appropriate agents for cooperation. We have proposed a new notion of *undistrust* and incorporated this, along with the notions of untrust and distrust proposed by Marsh [9], into the reasoning process. Our system is flexible; the fuzzy rules are specifiable by a system designer, and agents are able to scale inputs according to their current preferences regarding the relative importance of the trust dimensions. We have described initial experimental results that demonstrate the effectiveness of our approach in increasing the success rate, and the rate at which an agent's expectations are met in other trust dimensions. There are many areas of ongoing work, with our primary focus being additional experimentation to investigate different fuzzy rulesets and to consider the effect of different populations of reliable and unreliable agents. We also aim to integrate the model of individual fuzzy trust presented in this paper with existing models of reputation.

References

[1] C. Castelfranchi and R. Falcone. Principles of trust for MAS: Cognitive anatomy, social importance, and quantification. In *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS-98)*, pages 72–79, Paris, France, 1998.

[2] D. Gambetta. Can we trust trust? In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations*, pages 213–237. Basil Blackwell, 1988.

[3] N. Griffiths. Trust: Challenges and opportunities. *AgentLink News*, 19:9–11, 2005.
[4] N. Griffiths. Fuzzy trust for peer-to-peer systems. In *Proceedings of the P2P Data and Knowledge Sharing Workshop (P2P/DAKS 2006)*, to appear.
[5] N. Griffiths and K.-M. Chao. Experience-based trust: Enabling effective resource selection in a grid environment. In P. Herrman, V. Issarny, and S. Shiu, editors, *Proceedings of the Third International Conference on Trust Management (iTrust 2005)*, pages 240–255. Springer-Verlag, 2005.
[6] N. Luhmann. Familiarity, confidence, trust: Problems and alternatives. In D. Gambetta, editor, *Trust: Making and Breaking Cooperative Relations*, pages 94–107. Basil Blackwell, 1988.
[7] E. H. Mamdani and S. Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1):1–13, 1975.
[8] D. W. Manchala. E-commerce trust metrics and models. *IEEE Internet Computing*, 4(2):36–44, 2000.
[9] S. Marsh and M. R. Dibben. Trust, untrust, distrust and mistrust — an exploration of the dark(er) side. In P. Herrman, V. Issarny, and S. Shiu, editors, *Proceedings of the Third International Conference on Trust Management (iTrust 2005)*, pages 17–33. Springer-Verlag, 2005.
[10] NRC Institute for Information Technology. The FuzzyJ toolkit. www.iit.nrc.ca/IR_public/fuzzy/fuzzyJToolkit2.html, 2006.
[11] S. D. Ramchurn, C. Sierra, L. Godo, and N. R. Jennings. Devising a trust model for multi-agent interactions using confidence and reputation. *Artificial Intelligence*, 18(9–10):833–852, 2004.
[12] T. J. Ross. *Fuzzy Logic With Engineering Applications*. John Wiley & Sons, 2nd edition, 2004.
[13] J. Sabater and C. Sierra. REGRET: A reputation model for gregarious societies. In *Proceedings of the First International Joint Conference on Autonomous Agents in Multi-Agent Systems (AAMAS-02)*, pages 475–482, 2002.
[14] S. Song, K. Hwang, R. Zhou, and Y.-K. Kwok. Trusted P2P transactions with fuzzy reputation aggregation. *IEEE Internet Computing*, 9(6):24–34, 2005.
[15] N. Stakhanova, S. Basu, J. Wong, and O. Stakhanov. Trust framework for P2P networks using peer-profile based anomaly technique. In *Proceedings of the Second International Workshop on Security in Distributed Computing Systems*, pages 203–209, 2005.
[16] L. Xiong and L. Liu. PeerTrust: Supporting reputation-based trust in peer-to-peer communities. *IEEE Transactions on Knowledge and Data Engineering*, 16(7):843–857, 2004.
[17] L. A. Zadeh. A fuzzy-set-theoretic interpretation of linguistic hedges. *Journal of Cybernetics*, 2(3):4–34, 1972.