RELYING ON TRUST TO FIND RELIABLE INFORMATION

Alfarez Abdul-Rahman and Stephen Hailes
Department of Computer Science, University College London
Gower Street, London WC1E 6BT, United Kingdom
E-mail: {F.AbdulRahman,S.Hailes}@cs.ucl.ac.uk

Abstract

Information retrieved from open distributed systems can be of uncertain reliability and sifting through a large amount of data can be a complex procedure. In real life, we handle such problems through the social mechanisms of trust and word-of-mouth, or reputation. We propose a sociologically motivated trust and reputation model for the problem of reliable information retrieval and present an example of its use in this paper.

Keywords

Reputation; recommendation; trust; information retrieval; open distributed systems.

Introduction

We deal with the complexities and uncertainties of life by relying on two important social mechanisms – *trust* and *reputation*. However, our ability to use them is impaired when engaging in online interactions in large distributed systems such as the Internet. This is because there is currently no satisfactory trust model available to guide us through its complexity and to automate the sifting out of reliable information. Without a trust model, the task of identifying reputable and reliable information sources becomes daunting as *any* network user can be a source of information. This is the motivation underlying our work described in this paper. It is based on a more general trust model presented in (Abdul-Rahman and Hailes, 1997), which was designed to reason about trust in distributed systems security. Thus, this paper outlines a multi-disciplinary approach to the problem of information reliability, combining motivations rooted in distributed systems security with elementary foundations in sociology applied to the problem of reliable information retrieval. Indeed, such is the far-reaching nature of trust. The goal of this paper is to show that we can make information retrieval return more reliable results by applying the social mechanisms of trust and reputation.

Information Retrieval Problems

In general, increasing complexity and uncertainty are the main problems that users of the large and expanding Internet are faced with. Specifically, we will focus on three problem areas:

1. **Reliability of Retrieved Information:** Results of search engines return references to arbitrary pages of information containing the queried keyword(s). However they are unable to qualify the retrieved information with data about the sources of information themselves.

2. **Semantic mapping between agents[1]:** Keywords may sometimes have different meanings depending on context and even language (i.e. when a keyword is treated as simply a string of bytes). Furthermore, some keywords have subjective interpretations, like the words 'good' and 'beautiful'.
3. **Complexity Reduction:** It is simply beyond an average person's resources to filter through all information available from the Internet to select the ones best suited to his or her needs.

We argue that trust and reputation mechanisms can help reduce these problems in the following way. Socially, we tend to attach greater weight to opinions of people we know and trust (Hardin, 1991). Thus, by 'navigating' our information retrieval process through trusted sources, it is possible to obtain information that is more reliable than from arbitrary sources. This also automatically reduces the problem of information overload, since we only obtain information from sources we know we trust and can rely on. Thus, trust has also become a mechanism for complexity reduction (Luhmann, 1979). Furthermore, with the ability to handle reputational information in the Internet, a form of social control can be in place, clustering results towards those sources that are positively reputable (Rasmusson, 1996).

Trust

Trust is a social phenomenon. As such, any artificial model of trust must be based on how trust works in society. In this work, a survey of the social sciences was carried out, which highlighted the following properties of trust. We find it useful to use the following trust definition by Gambetta: "… *trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent will perform a particular action, both before [we] can monitor such action (or independently of his capacity of ever to be able to monitor it) and in a context in which it affects [our] own action*" (Gambetta, 1990).

Trust is not an objective property of an agent but a subjective *degree of belief* about specific agents (Misztal, 1996; McKnight and Chervany, 1996), within a specific context, that range from complete distrust to complete trust (Gambetta, 1990). However, trust is *not a prediction* nor some measure of probability (Luhmann 1979). A trusting action is taken despite *uncertainty* of outcome but in anticipation of a positive outcome (Baier, 1985; Misztal 1996; Barber, 1983). Nevertheless, that action may not follow the rules of rational choice theory (Barber, 1983; Gambetta, 1990) as an agent may have motivations other than risk and utility. It is also *non-monotonic* and the degree of trust is appraised with additional experience and information. Lastly, trust decisions are made based on the agent's knowledge and experiences (Hardin, 1993; Jøsang, 1996).

Reputation

In the words of Misztal (Misztal, 1996), "[Reputation] helps us to manage the complexity of social life by singling out trustworthy people – in whose interest it is to meet promises". In this work, a *reputation* is an expectation about an agent's behaviour based on information about or observations of its past behaviour.

Reputational information need not be solely the opinion of others. We also include reputational information completely based on an individual agent's own personal

[1] We will use the term *agent* to refer to any producer or consumer of information.

experiences. This allows us to generalise reputational information to combine personal opinions and opinions of others for the same reputation subject.

The Trust-Reputation Model

We propose a model for determining trustworthiness of agents based on the agent's collected statistics on 1) direct experiences and 2) recommendations about other agents from other agents (see Figure 1). Agents do not maintain a database of specific trust statements in the form of "*a* trusts *b* with respect to context *c*". Instead, at any given time, the trustworthiness of a particular agent is obtained by summarising the relevant subset of recorded experiences.

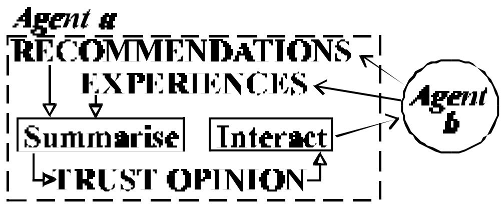

Figure 1 The Trust-Reputation Model

Direct and recommender Trust

The following represents an agent's belief in another agent's (*a*) trustworthiness within a certain context (*c*) to a certain degree (*td*) by the following:

$$t(a, c, td)$$

where $td \in \{vt, t, u, vu\}$. The semantics for *td* is given in Table 1 below. Additionally, we leave the context variable *c* open so that agents are able to define their own contexts when using this trust model.

vt	'Very Trustworthy'
t	'Trustworthy'
u	'Untrustworthy'
vu	'Very Untrustworthy'

Table 1 Trust degrees and their meaning.

An agent may also believe that another agent (*b*) is trustworthy to a certain degree (*rtd*) for giving *recommendations* about other agents[2] with respect to a context (*c*), represented as:

$$rt(b, c, rtd)$$

The value of *rtd* indicates the 'semantic closeness' between the recommendation and *x*'s own perception of the recommended agent's trustworthiness. For example, the recommender's perception of 'very trustworthy' may only equate to what *x* perceives to be 'trustworthy'. As with direct trust, we leave the context variable *c* open.

Experience Sets

An agent records its experiences with other agents in two separate sets: *Q* for direct trust experiences and *R* for recommender trust experiences. Assume that $C = \{c_1, \ldots c_n\}$ is the set

[2] We assume that recommenders may lie or give out contradictory recommendations to different agents.

of contexts known to an agent x and $A = \{a_1, \ldots, a_n\}$ is the set of agents that x has interacted with (either directly or as a recommender). An experience, e, is a member of the ordered set $E = \{vg, g, b, vb\}$, representing 'very good', 'good', 'bad' and 'very bad' respectively. These values correspond to the values given in Table 1. For each agent a and context c, there is an associated 4-tuple $s = (s_{vg}, s_g, s_b, s_{vb})$ where s_j is an *accumulator* for experiences $e = j$. Let $S = \{(s_{vg}, s_g, s_b, s_{vb})\}$. Q is defined as

$$Q \subseteq C \times A \times S$$

For experiences with recommender agents, the result is different. The goal is to obtain a *similarity measure* of an agent's recommendation and x's perception of the outcome. As a simple example, if a recommends to x that agent b is a 'very good' source of information with respect to context c, and x's evaluation of its experience with b is merely 'good', then future recommendations from a can be adjusted accordingly. In this example, we say that x's experience with b downgrades a's recommendation by one, or –1. The domain of possible adjustment values is given by the set $G = \{-3, -2, -1, 0, 1, 2, 3\}$.

There are 4 sets of adjustment experiences, T_{vg}, T_g, T_b and T_{vb}, for each recommending agent a and context c. Each T_e, where $e \in E$, represents adjustments for each of a's recommendations of e. The domain for values in T_e is the set G. Let $T = \{T_{vg}, T_g, T_b, T_{vb}\}$. R is defined as

$$R \subseteq C \times A \times T$$

Evaluating Direct Trust

To determine the direct trust degree td in an agent a with respect to context c, or in other words, "the reputation of a in context c", first obtain the relation (c, a, s) from Q. Let $s = (s_{vg}, s_g, s_b, s_{vb})$. Then the value of td is such that td is the subscript or index of s_e where s_e is the largest element in s.

$$\exists td \in E \ \ \forall s_e \in s \ , (s_e = \max(s)) \Rightarrow (td = e) \tag{1}$$

If max(s) returns more than one value, then td is assigned an *uncertainty* value according to Table 2 below (the symbol '?' indicates 'zero or one other value').

e	td	Meaning
$vg \wedge g \wedge ?$	μ^+	Mostly good and some bad.
$vb \wedge b \wedge ?$	μ^-	Mostly bad and some good.
All other combinations	μ^0	Equal amount of good and bad.

Table 2 Uncertainty Values.

Evaluating Recommender Trust

To determine the recommender trust degree rtd for an agent a in context c, we first find the relation $(c, a, t = (T_{vg}, T_g, T_b, T_{vb}))$ for a in R. The value of rtd is obtained by taking the **mod** of the *absolute* values of members in the set $T^a = T_{vg} \cup T_g \cup T_b \cup T_{vb}$. This tells how 'close' most recommendations are to the evaluating agent's own experiences by identifying how far an agent's recommendations usually are from the actual experience from relying on its recommendations.

$$rtd = \mathrm{mod}(\{\forall x \in T^a \mid |x|\}) \tag{2}$$

Evaluating Semantic Closeness of Recommenders

Let sc be a 4-tuple (sc_{vg}, sc_g, sc_b, sc_{vb}). To evaluate the 'semantic closeness', sc, of a recommender a in context c, first find the relation (c, a, t) in R, where $t = (T_{vg}, T_g, T_b, T_{vb})$. Then, for each member sc_e in sc, assign the **mod** of the corresponding member set T_e in t.

$$\forall e \in E, sc_e = \mathrm{mod}(T_e) \quad (3)$$

If T_e is multimodal, then this means that there is uncertainty in a's recommendations and further experience is required to resolve this uncertainty. In this case, we let $s_e = 0$ so that no adjustments are made to a's recommendations. This allows us to take future recommendations at 'face value' and decide on the difference after the experience of relying on those uncertain recommendations.

Evaluating a Recommendation

To evaluate a recommendation of degree d from a about b's trustworthiness in context c, represented by $rec(a, b, c, rd)$, where $rd \in E$, first evaluate the semantic closeness, $sc = (sc_{vg}$, sc_g, sc_b, $sc_{vb})$, of a for context c as shown in the previous section. Then adjust rd using the appropriate sc member to obtain the adjusted recommended trust degree, rd^*, shown as

$$rd^* = rd \oplus sc_{rd} \quad (4)$$

where $\oplus$ denotes the operation 'is increased by the order of'. E.g., if $rd = vg$ and $sc_{vg} = -1$ (i.e. downgrade by one) then $rd^* = vg \oplus -1 = g$.

Updating Experiences

After an experience with an agent a, the experience relation for a in Q is updated by incrementing the appropriate experience type counter. For example, if (c, a, $s = (s_{vg}$, s_g, s_b, $s_{vb})$) is the appropriate relation in Q, and it was a 'good' experience (e = g), then increment s_g. Formally, given an experience of e,

$$s_e = s_e + 1 \quad (5)$$

Furthermore, if the experience was a result of relying on a recommendation from agent b, then we also update the experience for b in R by obtaining the appropriate relation in R for b, (c, b, t), where $t = \{T_{vg}, T_g, T_b, T_{vb}\}$ and adding the difference between the recommended trust degree, rd, and the experience, e, to T_{rd}. The 'difference', shown by the operator $\lozenge$, is the number of levels to upgrade or downgrade rv to get e, e.g. if rd = 'very good' and e = 'good' then $e \lozenge rd = -1$, or 'downgrade by one'.

$$T_{rd} = T_{rd} \cup \{e \lozenge rd\} \quad (6)$$

Combining Recommendations

Sometimes an agent x may encounter more than one recommender for a particular recommended agent y. To evaluate the final trust degree of y, ct_y, by combining the recommendations, we first obtain the recommender trust value, for all *known* recommenders of y (recommendations from unknown agents are discarded). If $a_1 \ldots a_n$ are the recommenders, then obtain rtd_k in each $rt(c, a_k, rtd_k)$ for $k = 1 \ldots n$. Each recommender is then assigned a weight according to its rtd_k value using Table 3 below:

rtd_k	0	1	2	3	unknown
weight[3]	9	5	3	1	0

Table 3 Recommender weights.

We then adjust the recommendations according to (4). Now, for each recommended trust degree $e \in E$, sum the weightings of the recommenders who recommended e. Assume $a_1 \ldots a_n$ are recommenders of y, $w_1 \ldots w_n$ are their corresponding individual weightings (i.e. w_n is the weighting for a_n) and $rec(a_n, b, c, rd_n)$ is a_n's recommendation. Let L_e be the set whose members are the weights associated with recommenders who recommended e, then

$$\forall e \in E \ \ \forall w_i \in L_e, \ sum_e = \sum_{i=1}^{|L_e|} w_i \tag{7}$$

ct_y is the sum_e with the highest value. If there are more than one largest sum_e, then ct_y is assigned an uncertainty value according to Table 2.

Conclusion

We have discussed why trust and reputation is important for reliable information retrieval and proposed a model for trust and reputation motivated by their 'real-world' characteristics as identified in the social sciences. However we acknowledge that the approaches has some ad-hoc design decisions, namely the members of E and the weightings. Future work will include further investigation into a more concrete basis for these values.

References

Abdul-Rahman, Alfarez and Hailes, Stephen (1997); A Distributed Trust Model; Proceedings of New Security Paradigms 97 Workshop

Baier, Annette (1985); Trust and Antitrust; Ethics, 96 (pp. 231-260)

Barber, Bernard (1983); The Logic and Limits of Trust; New Brunswick, New Jersey

Gambetta, Diego (1990) Can We Trust Trust?; In Trust: Making and Breaking Cooperative Relations, Gambetta, D (ed.), Basil Blackwell, Oxford, (pp. 213-237)

Hardin, Russell (1993); The Street Level Epistemology of Trust; Politics and Society, 21 (pp. 505-531)

Jøsang, Audun (1996); The right type of trust for distributed systems; Proceedings of New Security Paradigms 96 Workshop

Luhmann, Niklas (1979) Trust and Power; Wiley, Chichester

McKnight, D. Harrison and Chervany, Norman L. (1996); The Meanings of Trust; Technical Report 94-04, Carlson School of Management, University of Minnesota

Misztal, Barbara (1996); Trust in Modern Societies; Polity Press, Cambridge MA

Rasmusson, Lars (1996); Socially Controlled Global Agent Systems; Master's Thesis, Swedish Institute of Computer Science

[3] The weightings were chosen almost arbitrarily, but in such a way as to reflect the relative importance of recommendations from agent who are semantically closer.