Department of Mathematics
Faculty of Mechanical Engineering
Slovak University of Technology in Bratislava
Aplimat
4th International Conference
APLIMAT 2005

HUMAN TRUST MODELLING

NETRVALOVÁ Arnoštka (ČR), ŠAFAŘÍK Jiří (ČR)

Abstract. The paper deals with the human trust modelling. The terms trust and trust representation and visualization are discussed. Our approach in human trust modelling is based on the theory of information and social communication knowledge. Some required terms from the agent theory are mentioned. Agent approach is chosen for modelling the trust in the community. Fundamentals of trust formation, trust dissemination, and trust evolution are presented deploying the agent system.

Key words. Trust, trust modelling, information dissemination, agent system

1 Introduction

There are already many studies coming from psychological or social sciences there are examining the meaning and characteristics of trust. Most of these works deal with examination of the trust behavioural pattern using computational simulation. The aim of our future work will be modelling and simulation of the trust evolution. The possible appropriate concepts are described in next sections.

2 Trust, trust representation and visualization

Some necessary terms are discussed to easier understanding of context of trust modelling. The acceptation of trust is wide. The World Book Dictionary [12] offers further explanations, as: firm belief in the honesty, truthfulness, justice, or power of person or thing**;** a person or thing trusted; confident expectation or hope**;** something managed for the benefit of another; something committed to one's care**;** the obligation or responsibility imposed on one in whom confidence or authority is placed; condition of one in whom trust has been placed; confidence in the ability or intention of a person to pay at some future time for goods or services; business credit. To summarize, we will understand the trust as *given credit, hope, and confidence in the ability or intention of a person to pay at some future time for service* for the purpose of building models of it.

Let we think of some phrases, e.g. “I trust him.”, or “He trusts them.”, etc. What does it really mean? Can be trust measured? These questions cross one’s mind. These questions can be answered by using some simplifications and limited presumptions. For examining the trust as a behavioural pattern, we need some ways of its representing and visualizing. It is possible

to create some methods that can measure and visualize the trust. Generally, we can say that behaviour of entities has bounded rationality. The behaviour of an entity can be a state of its mind, unexpected event in its surroundings, etc. For a distant observer, it can be a random event that causes more or less unpredicted behaviour. The most straightforward way to model random events is to use random numbers.

By modifying Marsh's approach [4], we treat the trust as a value between 0 and 1, where 0 means “complete distrust” and 1 means "blind trust". We may illustrate this concept as shown in Figure 1. Another way to treat the trust is using values between -1 and 1, where -1 is representation of complete distrust and 1 represents blind trust.

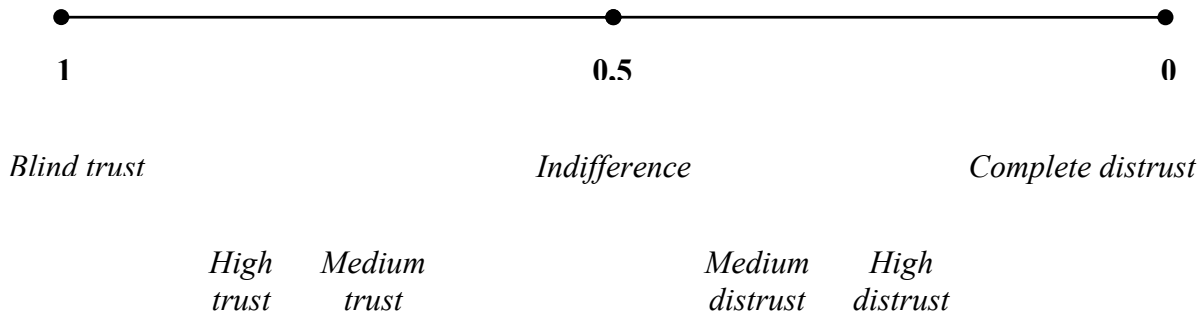

Figure 1 Representation of trust - trust value is visualized as a point on the interval <0, 1>

The trust is usually or shared between two or among more entities in some community. Then, we may say that it is a property of the relationship between entities. For the reason of simplicity, let us consider the community of entities to be composed from the couples of single objects. Let us consider a couple with two relationships and one trust value per relationship. In the work [8] these trust values are denominated T_L and T_R meaning trust from left to right and from right to left respectively. A square can be drawn in a two dimensional coordinate system (Figure 2). The trust values T_L, T_R are projected onto the two perpendicular sides of the square. Thus, the trust between the objects of the couple may be treated as a two-dimensional vector (T_L, T_R). It is a point in the square, thus visualizing the trust in the couple. This is shown in the Figure 2. It is very simple visualization; therefore it is easily and quickly readable.

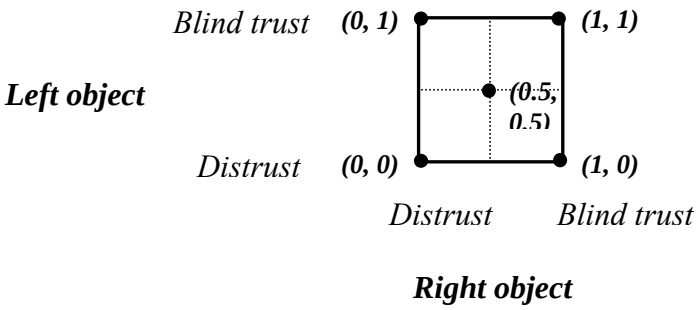

Figure 2 Trust square - simple trust visualization of couple of objects

Now, we will discuss the trust visualization. When representing the point with coordinates (T_L, T_R) by a small black circle in the trust square, we get different shapes of the trust square. Trust square shapes can be simple visualization of trust rate in the couple. Some of basic squares are shown in Figure 3. Shape 1 denotes a couple with mutual distrust. Shape 2 represents couples where an entity trusts the other one and the other entity distrusts completely. Shape 3 shows situation, where an entity trusts and the other one is indifferent. The opposite situation is shown in the shape 4 where an entity is indifferent and distrusts the other one. The shape 5 denotes that both entities are indifferent to each other, or that there are no relationships between them. Taking in the account our simplifying assumption, the set of trust squares visualizes the trust in the community of entities.

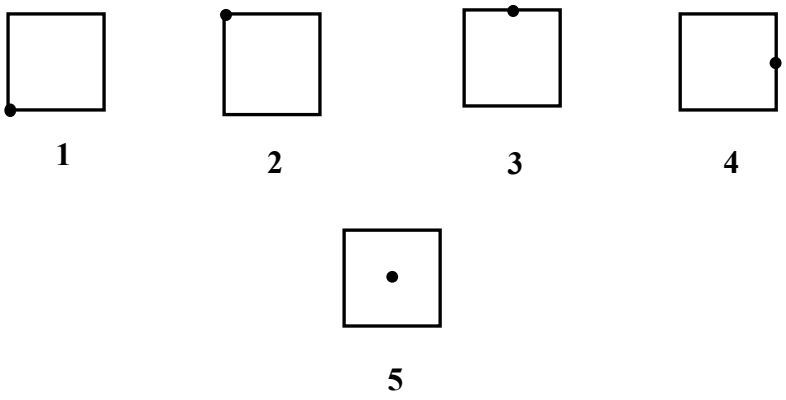

Figure 3 Some of basic trust square shapes - simple level of trust visualization

2 Application theory of information to trust, a basic

If we want to trust to some one (or something), we have had some information. How to describe the information as the measurable value? We can use the probabilistic and statistic approaches. The concept of information is too broad to be captured completely by a single definition [7]. The notion of entropy has many properties that agree with the intuitive notion of what a measure of information should be. This notion is extended to define mutual information, which is a measure of the amount of information one random variable contains about another. Entropy then becomes the self-information of a random variable. Mutual information is a special case of a more general quantity called relative entropy, which is a measure of the distance between two probability distributions. All these quantities are closely related and share a number of properties. We present some of these properties.

The entropy $H(X)$ of a discrete random variable X with the probability mass function $p(x)$ is defined by

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \tag{1}$$

Now, we extend the entropy definition to a pair of random variables. The joint entropy $H(X, Y)$ of a pair of discrete random variables (X, Y) with a joint distribution $p(x, y)$ is define as

$$H(X,Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x,y) \tag{2}$$

We also define the conditional entropy of a random variable given as the expected value of the entropies of the conditional distributions averaged over the conditioning random variable. If $(X, Y) \sim p(x, y)$, the conditional entropy is defined as

$$H(Y \mid X) = \sum_{x \in X} p(x)\, H(Y \mid X = x) = -\sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(y \mid x) \tag{3}$$

Now, we describe the relative entropy which is a measure of the distance between two distributions. The relative entropy $D(p \,||\, q)$ (or divergence) between two probability mass functions $p(x)$ and $q(x)$ is defined as

$$D(p \,||\, q) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)} = E_p \log \frac{p(X)}{q(X)} \tag{4}$$

Further, we introduce mutual information, which is a measure of the amount of information that one random variable contents about another random variable. Consider two random variables X and Y with a joint probability mass function $p(x, y)$ and marginal probability mass functions $p(x)$ and $p(y)$. The mutual information $I(X; Y)$ is the relative entropy between the joint distribution and the product distribution $p(x)p(y)$, i.e.,

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = D(p(X, Y) \,||\, p(X)p(Y)) = E_{p(x,y)} \log \frac{p(X, Y)}{p(X)p(Y)} \tag{5}$$

Note that $D(p \,||\, q) \neq D(q \,||\, p)$ in general.

By rewriting the definition of mutual information $I(\mathrm{X}; Y)$ the relationship between entropy and mutual information is defined as

$$I(X;Y) = \sum_{x,y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} = H(X) - H(X|Y) \tag{6}$$

Thus, the mutual information $I(X; Y)$ is the reduction in the uncertainty of X due to the knowledge of Y and vice versa. More on this topic is in [2]. Some generalization of this traditional theory is in [9, 3].

Here is presented situation in which model and reality are different. Comparison is in Table 1 below. Heading distribution will be described as $p(x)$, its model (estimation) $e(x)$ and comparative probability $q(x)$.

It's possible to use $p(x)$ as a model of temporary situation and $e(x)$ as a model of new state after spreading some message, e.g. as a model trust dissemination. The second case is interesting for purposes of this paper.

Table 1 Shannon's classical theory in comparison with the concept of disinformation

Measure	Shannon's, classical	Concept of disinformation
Entropy	$H(X) = -\sum_x p(x) \log p(x)$	$H(X;e) = -\sum_x p(x) \log e(x)$
Mutual information	$I(X:Y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$	$I(X:Y;e) = \sum_x \sum_y p(x,y) \log \frac{e(x,y)}{e(x)e(y)}$
Divergence of probability models	$D(p \parallel q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$	$D(p \parallel q;e) = \sum_x p(x) \log \frac{e(x)}{q(x)}$
Symmetrical divergence of probability models	$J(p \parallel q) = \sum_x (p(x) - q(x)) \log \frac{p(x)}{q(x)}$ $J(p \parallel q) = D(p \parallel q) + D(q \parallel p)$	$J(p \parallel q;e) = \sum_x (p(x) - q(x)) \log \frac{e(x)}{q(x)}$ $J(p \parallel q;e) = D(p \parallel q;e) + D(q \parallel e)$

4 Aspects of trust modelling in agent system

In section 2, we considered the trust in the community of entities. It is rather straightforward to model the trust in the population using agent systems [10].

We can think of agents as of living entities. Our agents are traditional agents with memory, energy supply, receptors and effectors. They have ability to observe, act, remember, reproduce and die. Agent's energy supply is a simplified concept of the life energy. Basically, the energy is used for performing agent actions, including reduction. By running out of the energy, agent dies. Memory is agent's organ that has an ability to collect, store and forget observed information. The agent system is characterized by environment, where the agent operates. The agent is autonomous unit that is furnished by certain ration and is able to solve some specific problems [5, 6]. The result of agent action is the transition of an agent system from initial to required state.

Agents can influence the behaviour of each other. They can maintain their actions with the others, or not disturb the actions the others, or even act against the others. Thus, we can divide the agents into cooperative agents that have joint intentions, or competitive agents that have antagonistic intentions, and collaborative agents that cooperate with each other.

Agent's strategy describes which action will be done as an actual status reaction to environment. Dominant strategy is such that it is the best individual strategy without seeing strategy of the others [11]. The rational agent votes always dominant strategy. Strategy of group is Nash equilibrium, which describes that each of strategy is the best individual strategy of competent agent due to selected strategies of other agents. Generally, the strategy choice leading to optimal acquisition of the whole group is exigent of coordination the negotiation of all agents. They must communicate to each other and need the will to benefit of whole group. Agent's group can have joint mental poses pronounce by formulas and all of agents must know of them.

Common mental poses are the background for making agreements and coalitions. The agents that create a group accept the commitments and general rules, and abide by these norms. Collaborating agents must have the capability of communication to each other. It enables the coordination of their actions and searching the joint strategies for the goal acquirement in joint

interest. Negotiation is a technique for reaching an agreement on a matter of mutual interest. The agents must have nested the basic rules collaboration in the knowledge bases. Furthermore, the agents may be able to plan their activity. Rational agent comes in the collaboration commitment with other agents when it may promise some profit only. The agents reach by agreement better environment status, than they have reached with autonomous non-coordinated action, or they reach the compromise in the course of the interest conflict. Collaboration and compromise bears on share goals, resources and interest conflicts. It calls into existence the cooperation agreements and the conflict agreements. The agents express their will to collaboration by way of commitments. The commitment is the maintenance of the mental pose. Mostly, the agents concert the conditional agreements, but only the standards are exception. The agents must hold the standards for all the time of their existence. We may categorize agent groups by the reason of collaboration interests on the groups that share goals, or share resources, or furnish information.

We intend to introduce the trust in community of agents. The rules deploying the measure of trust described in sections 2 and 3 will be proposed.

5 Formation, dissemination and evolution of trust

Agent approach is chosen for describing formation, dissemination and evolution of trust. The modern agent technology is used to promote non-trivial interactions among agents and reduce risk transactions as much as possible. The development of trust-based collaborations is necessary. This requires some trust management framework (TMF) that enables to form, maintain and evolve trust opinions.

We give an overview of the Trust Management Model that was developed by Capra, for details refer to [1]. The structure of the model is shown in Figure 4. Three components create the model. They are Trust formation, Trust dissemination and Trust evolution. If the agent *a*, which is called the "trustor", has decided for another agent *b* which is called "trustee" trust information about agent *b* has to be collected. The experience represents the history of agent and it is saved in the local environment. The recommendations from other agents are propagated by means of the Trust dissemination component. Trust information is processed by the Trust formation component. These facts help to predict the trustee's trustworthiness.

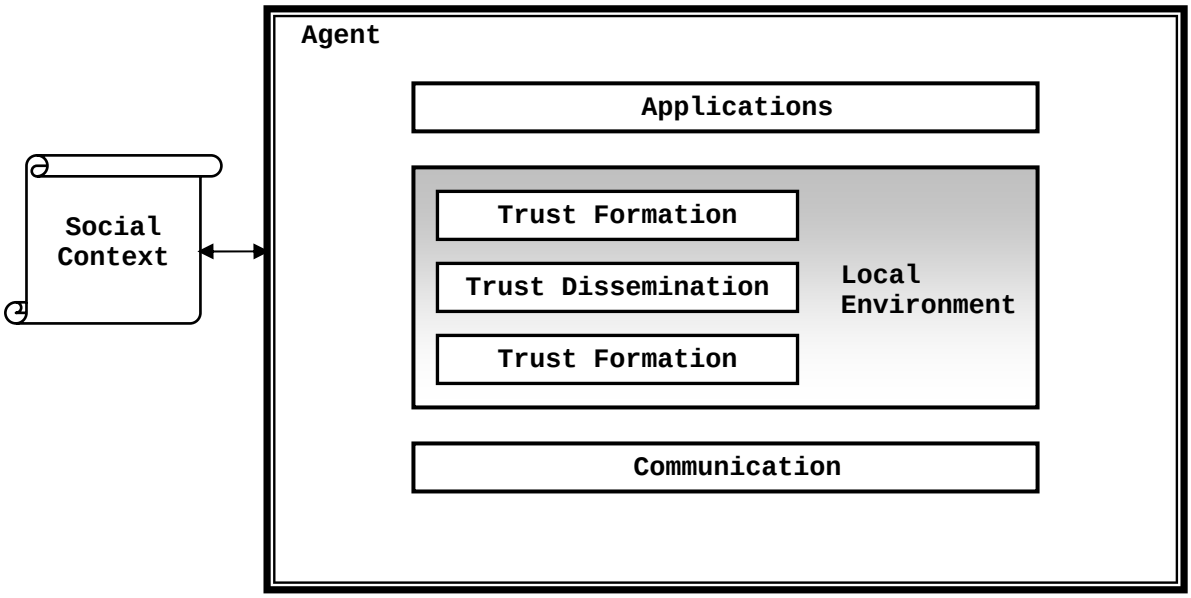

Figure 4 Trust Management Model

The process that enables a trustor agent to predict a trustee's trustworthiness before the interaction takes place is called trust formation. There is need to have information that it is used to predict trustworthiness and the trust formation function that is used to compute a prediction. A trustor forms a trust opinion about a trustee by aggregated trust information. Aggregated trust information is the information locally kept by Trust Management Framework (TMF) and mainly based on past direct experience with other agent (from their transactional context) and recommendations that sent to the agent by others in the social context (only those that in the past interacted with).

The trust formation function uses recommendations to predict trustworthiness of a trustee. These recommendations are important when the trustee is unknown to the trustor. Then some protocol for dissemination of recommendation is indispensable. This recommendation exchange protocol guarantees trustor minimum set information for create the prediction. Each trustee carries a portfolio of credentials. The portfolio is a set of letters of presentation that represent the history of the agent itself. Each letter comprises information and it is authentic with his private key. Thus, in the exchange protocol the trust information function is computed in three different events: prior to its execution, after to obtain portfolio of credentials and once again if further recommendations are received form the social context. The trustworthiness of trustee is based on past experiences as perceived by trustor.

The evolution is the continuous self-adaptation of trust information that is kept in the local environment of agent. There is need to introduce two further functions. These functions are an aggregation function and a tacit information extraction. The aggregation function is used to update the perceived trustworthiness of trustee when a new direct experience between two agents occurs. Only if there is no interaction then the trustworthiness of trustee may be updated. It is based on the recommendations received about trustee from trusted recommenders. Thus, the trust information for each agent is minimal. The aggregated information is signed by private key of trustor. It is used to provide the trustee a letter of presentation at the end of the exchange protocol. Likewise it is used to answer request for recommendations that come from other agents in social context. There is known when trustor has to make a trust decision about trustee (without previous direct experiences) there are only recommendations to rely on. Because the trust is subjective, these recommendations can be conflicting with each other. In this case, the weighing of more recommendations is used. The recommendation coming from an agent with whom there is no share of experience has weighing less or it is even discarded. The tacit information contains information about trustworthiness of agent as recommender. This information is updated based on the perceived trustworthiness of trustee with whom trustor has just interacted and the recommendation about trustee. The both of functions adjust the value of an agent's trustworthiness based on behaviour. The trust is changed dynamically when behaving is well and loses when there is misbehaving. It was observed that the more accurate the agent's knowledge of the surroundings becomes, the more frequently the agent has interacted, and conversely.

6 Conclusions

We have described two approaches to the measurement of the trust and outlined their deployment for the modelling of the trust in the community. The existing model of the trust management using agent technology was considered as possible environment for exploring the different measures of the trust.

References

[1] Capra, L.: *Engineering Human Trust in Mobile System Collaborations*, Dept. of Computer Science, University College London, UK, 2001

[2] Cover, T. M., Thomas, J. A.: *Elements of Information Theory*. Wiley, 1991.

[3] Kotlíková, M., Mašková, H., Netrvalová, A., Nový, P., Spíralová, D., Vávra, F., Zmrhal, D.: *Informace a dezinformace - statistický pohled*. Letní škola JČMF ROBUST, Třešť, 2004

[4] Marsh, S.: *Formalising Trust as a Computational Concept*, Ph.D. Thesis, Department of Mathematics and Computer Science, University of Stirling, 1994

[5] Mařík V.: *MAS-řešení složitých a rozsáhlých úloh*, přednášky Umělá inteligence, FEL ČVUT, Praha, 2002

[6] Pěchouček, M.: *Introduction to Multi-Agent Systems*, Agent Technology Group, Gerstner Laboratory, Department of Cybernetics, Czech Technical University, 2002

[7] Po-Ning Ch., Fady A.: *Lecture Notes in Information Theory*, Vol. I, department of Communication Engineering, National Chiao Tung University, Taiwan, Republic of China; Department of Mathematics & Statistics, Queen's University, Kingston, Canada, 2004

[8] Urbánek, Š.: *Modelling and Simulation of Trust Evolution in Complex Systems*, Master Thesis, STU - Faculty of Informatics and Information Technologies, Bratislava, 2004

[9] Vávra, F., Nový, P.: *Informace a dezinformace*. Seminář z aplikované matematiky. Katedra aplikované matematiky, Přírodovědecká fakulta MU, Brno, 2004

[10] Wooldridge, M., Jennings, N.: *Intelligent Agents: Theory and Practice*, Knowledge Engineering review, 1995

[11] Zbořil F.: *Plánování a komunikace v multiagentních systémech*, Disertační práce, Fakulta informačních technologií, VUT Brno, 2004

[12] The World Book Dictionary, World Book, Inc. a Scott Fetzer company, The World Book Encyclopaedia, Chicago 1988

Contact address

Ing. Arnoštka Netrvalová, netrvalo@kiv.zcu.cz
Prof. Ing. Jiří Šafařík, CSc., safarikj@kiv.zcu.cz

Department of Computer Science and Engineering,
Faculty of Applied Sciences,
University of West Bohemia,
Univerzitní 22, 306 14 Plzeň,
Czech Republic