A Versatile Approach to Combining Trust Values for Making Binary Decisions[1]

Tomas Klos[a] Han La Poutré[a,b]

[a] *Center for Mathematics and Computer Science (CWI), Amsterdam*
[b] *Faculty of Technology Management, Technical University Eindhoven*

1 Introduction

In open multi-agent systems such as grid-based virtual organizations [1], agents typically need to rely on others for the delivery of information or resources or for the execution of tasks. Since trustworthiness can not be taken for granted, however, an agent needs to build up a measure of her trust in other agents in her environment, and to update it on the basis of her experiences with those other agents.

Many different computational trust models have been proposed in the literature, based on a wide variety of techniques (see, e.g., [2] for a review of computational trust models in multi-agent systems). In the current paper, we let the agents use Bayes' rule to update beliefs about other agents' capabilities. Once an agent has established a trust or reputation value for other agents, resources, services, etc., the agent needs to act *on the basis of* those values. This typically involves estimating the value of a certain, often binary, random variable. If different agents are providing conflicting information about this value, and some are trusted (highly) and some are not, then how should one combine these trust values with the information provided? Which of the agents are more important, and how should conflicting claims be weighted? We investigate the relative effectiveness of a variety of methods for combining trust values.

2 Establishing and Combining Trust Values

For establishing trust values, we let agent i assume that each other agent j's behavior ($j = 1, \ldots, J$) consists of a sequence of Bernouilli trials with 'success' and 'failure' as possible outcomes, and is governed by an agent-specific parameter θ_j: the 'bias,' or probability of success. Agent i's trust in agent j refers to i's beliefs about the value of θ_j, which i updates using Bayes' rule. When i uses the Beta(a, b) distribution as a (conjugate) prior distribution for θ_j, then the posterior distribution given our binomial likelihood is also a Beta distribution. Trust is then usually taken as the expected value of this posterior Beta distribution which is easily expressed using its parameters a and b as $E[\theta_j] = \frac{a}{a+b}$. Updating is then simply accomplished by taking $a = u + 1$ and $b = v + 1$, where u and v are the previous positive and negative experiences i has had with j, and where the values of 1 are the parameters of the Beta$(1, 1)$ distribution, the uniform prior. Some example Beta distributions with different numbers of positive and negative experiences are shown in Fig. 1(a), with the corresponding expected values depicted using vertical lines.

We propose 3 methods for combining trust values when the observations of the value $a \in \{0, 1\}$ of a binary variable A, by all J observers in the collection $\mathcal{J}$ need to be combined.

majority (m) This method lets agent i simply report the value for A which is observed by the majority of agents in $\mathcal{J}$. (We only allow odd values for J.)

[1]This paper was originally published in Ketil Stølen, William H. Winsborough, Fabio Martinelli, and Fabio Massacci, editors, *Proceedings 4th International Conference on Trust Management, May 16-19 2006, Pisa, Italy*, volume 3986 of Lecture Notes in Computer Science. Springer, 2006.

evidence (e) Agent i adds her positive and negative experiences across all agents claiming each of the values $a \in \{0,1\}$ and estimates the expected capability in each of the 2 groups of agents.

$$E(\theta_{\mathcal{J}_a}) = \frac{\sum_{j \in \mathcal{J}_a} \text{positive}_i^j + 1}{\sum_{j \in \mathcal{J}_a} \text{positive}_i^j + \sum_{j \in \mathcal{J}_a} \text{negative}_i^j + 2}$$

She chooses the value for a which maximizes $E(\theta_{\mathcal{J}_a})$.

likelihood (l) Using this method, assuming each value in turn to be correct, agent i calculates the joint probability of the observations, which is the same as the likelihood of each of the two values for a, given the observations. For each value of $a \in \{0,1\}$ (where $\bar{a} = 1 - a$), the likelihood of it being the true value of A is equal to

$$L(A = a|\text{observations}) = p(\text{observations}|A = a) = \prod_{j \in \mathcal{J}_a} \text{trust}_i^j \prod_{j \in \mathcal{J}_{\bar{a}}} (1 - \text{trust}_i^j)$$

where $\mathcal{J}_b$ (for $b \in \{a, \bar{a}\}$) is the subset of $\mathcal{J}$ claiming $A = b$. Agent i reports the value $b \in \{a, \bar{a}\}$ which maximizes the likelihood $L(A = b|\text{observations})$.

3 Results

Figure 1(b) shows a sample of our results. In this case, all $J = 7$ observers have the same probability of

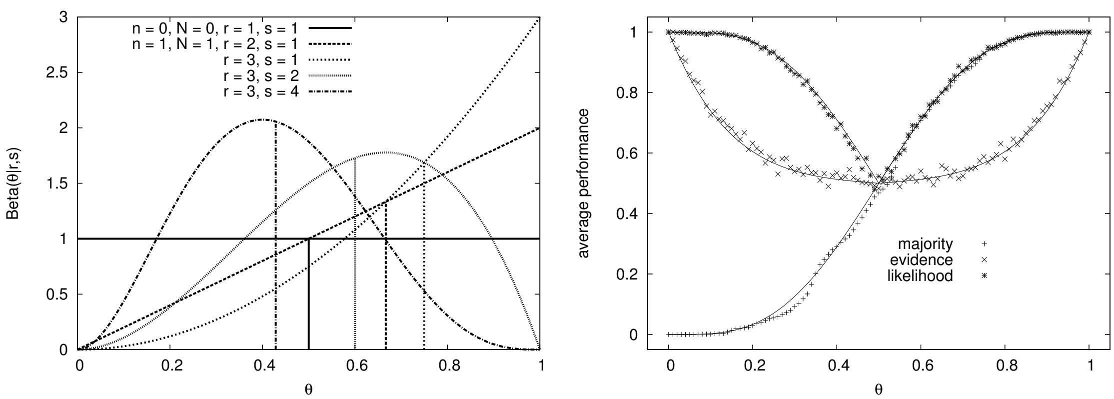

Figure 1: 1(a): example Beta distributions; 1(b): performance of the methods.

success, θ_j, shown on the x-axis. The graph shows the performance at time $t = 1000$ of agent i using each of the 3 methods, averaged over 1000 replications of each experiment (the crosses and plusses). In this simple case, each method's expected performance can be calculated analytically (the solid lines). More interesting results arise when different values for θ_j are distributed less evenly across the different observers. The various methods show differential sensitivity to such changes, but the likelihood method is typically optimal.

References

[1] Foster, I., Jennings, N.R., Kesselman, C.: Brain meets brawn: Why grid and agents need each other. In: Proc. AAMAS 2004, ACM (2004) 8–15

[2] Sabater-Mir, J., Sierra, C.: Review on computational trust and reputation models. AI Review **24** (2005) 33–60