ABSTRACT
The Semantic Web, conceived as a collection of agents, brings new opportunities and challenges to trust research. Enabling trust to ensure more effective and efficient interaction is at the heart of the Semantic Web vision. We propose a composite trust model based on statistical decision theory and Bayesian sequential analysis to balance the costs and benefits during the process of trust evaluating and combine a variety of sources of information to assist users with making the correct decisions in selecting the appropriate service providers according to their preferences. The model proposed by this paper gives trust a strict mathematical interpretation in terms of probability theory.

Categories and Subject Descriptors
I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence – Intelligent agents, Multiagent systems; F.1.2 [Computation by Abstract Devices]: Modes of Computation – Probabilistic computation.

General Terms

Keywords
Trust Model, Semantic Web, Bayesian Sequential Analysis.

1. INTRODUCTION
In our model, the quality of a provider agent can be considered to be an unknown numerical quantity, and will represent it by \( \theta \) and it is possible to treat \( \theta \) as a random quantity with a probability distribution. Consider the situation of an agent \( A \) try to make an estimate of trust value for provider agent \( B \). \( A \) holds a prior information (subjective) of \( B \), represented by distribution \( \pi(\theta) \), and request \( A \)'s acquaintance to give opinions on \( B \)'s quality. After \( A \) receives the assessments of \( B \)'s quality from his acquaintances, \( A \) takes these statements as sample about \( \theta \). Outcome of these sample is a random variable and will be denoted \( X \). A particular realization of \( X \) will be denoted \( x \) and \( X \) will be assumed to be a random variable, with density \( f(x|\theta) \). Then, we can compute "posterior distribution" of \( \theta \) given \( x \), denoted \( \pi(x|\theta) \). \( \pi(x|\theta) \) reflects the update beliefs about \( \theta \) after (posterior to) observing the sample \( x \). We take the posterior distribution of \( \theta \), \( \pi(x|\theta) \), as the estimate of \( B \)'s quality. If we want to take further investigation on \( B \)'s quality for more accuracy, \( \pi(x|\theta) \) will be used as prior distribution for the next stage of investigation instead of original \( \pi(\theta) \).

When several provider agents exist, \( A \) need to decide which one should be selected. At that time, the preferences of agent \( A \)'s owner should be considered properly to make this decision. Therefore, utility function should be constructed for agent \( A \)'s owner, which represented by \( U_A(x) \), to express his preferences, where \( r \) represents rewards of the consequences of a decision. Supposing that \( \pi(x|\theta) \) is the posterior distribution of agent \( B \), the expected utility of function \( U_A(x) \) over \( \pi(x|\theta) \), denoted \( E^{\pi(x|\theta)}[U_A(r)] \), is possible gain of consequence of selecting \( B \).

2. COMPOSITE TRUST MODEL
By treating an agent as a node, the "knows" relationship as an edge, a directed graph emerges. To facilitate the model description, agents and their environment are to be defined. To clarify the idea of our trust model, we begin with a simple illustration. Consider the scenario that agent \( A \) is evaluating trust value of \( B \) and \( C \) for being business. The set of all consultant agents that \( A \) requests for this evaluation as well as \( A \), \( B \), \( C \) can be considered to be a unique society of agents. In our example (see Figure 1) it is called a "closed society of agents" with respect to \( A \).

2.1 Closed Trust Model
We will treat the service quality, \( \theta \), as a continuous variable here, and suppose that the service charges of \( B \) and \( C \) are 250 and 210 units respectively \((SC_B = 250, SC_C = 210)\), where \( SC \) denotes service charge). As shown in Figure 1, the agent \( A \) feels that \( \theta_B \), the service quality of \( B \), has a normal prior density, \( N(0.64, (0.08)^2) \). We also suppose that the prior density of \( \theta_C \) is \( N(0.5, (0.15)^2) \) here. The probability distribution of \( X \) that represents the assessments of service quality from consultant agents will, of course, depend upon the unknown state of nature \( \theta \). Therefore, we assume that \( A \) is another continuous random variable with density \( f(x|\theta) \sim N(\theta, (0.05)^2) \). We also assume that the agent \( A \)'s owner has a utility function \( U_A(r) = 208.4 + 459.8x + 99.84r^3 \).
The sample from the agent with shorter referral distance should be used with higher priority, we propose following order rules: (1) to answer the question of which sample information should be prioritized, (2) to answer the question of whether the sample information with smaller reliability factor should be used before the one with larger reliability factor is used, hence (225.34 > 195.82), (3) to answer the question of whether the posterior distribution of one agent is prior to that with smaller one.

Now, we can evaluate trust value of to that with larger reliability factor is prior to that with smaller one.

We take the assessments of B's quality from the consultant agents as sample about to and combine these sample information (x) and prior information into posterior distribution of given x. In order to answer the question of which sample information should be used with higher priority, we propose following order rules: (1) The sample from the agent with shorter referral distance should be used first; (2) If the samples come from the agents that have the same referral distance, which with larger reliability factor is prior to that with smaller one.

Note that the marginal distribution of, is known. Hence, after rectification, is unknown but is known. Let be prior information, , a N(m, r²) density, where and r² are known. Note that the marginal distribution of, is N(m, r² + r²) and the posterior distribution r of give x is N(ρ, r²), where

\[ \mu(x) = \frac{\sigma^2 + r^2}{\tau^2 + \sigma^2 + r^2} x + \frac{\tau^2}{\tau^2 + \sigma^2 + r^2} \]

We can evaluate trust value of B for A by using above formulas and information. Firstly, we use sample information from D through the path A→D→B, and the posterior distribution \( \pi_d(\theta|x = 0.55) \sim N(0.5753, (0.0424)^2) \), where

\[
\begin{align*}
0.05^2 \times 0.64 + 0.08^2 \times 0.55 &= 0.55753 \\
0.05^2 \times 0.55 &= 0.02412 \\
0.08^2 \times 0.55 &= 0.04241
\end{align*}
\]

However, the reliability factor has not been considered during the process of calculating this posterior distribution. We used following formula to rectify above \( \pi_d(\theta|x = 0.55) \), where \( p_{old} \) and \( p_{new} \) represent the parameter of prior distribution, the posterior distribution before rectification and the posterior distribution after rectification respectively (p is mean or variant for normal distribution), and R is reliability factor.

\[
p_{new} = p_{old} + (p - p_{old}) \times R
\]

Hence, after rectification, \( \mu_B = 0.64 + (0.5753 - 0.64) \times 0.82 = 0.5869 \) and \( \tau_B = 0.08 + (0.0424 - 0.08) \times 0.82 = 0.0492 \). Then the posterior distribution \( \pi_d(\theta|x = 0.55) \sim N(0.5869, (0.0492)^2) \). The residual calculating process of B and the whole process of C are shown in Table 1 and 2. Note that we employ multiplying to merge two or more than two reliability factors. The reason behind using multiplying is that if the statement is true only if the agents who propagate this statement all tell the truth and it is considered to be independent for any two agents to lie or not to lie. Expected utility of B is:

\[
\begin{align*}
\text{utility of } B &= \int_{-\infty}^{\infty} U_A(r) \pi_B(\theta | x) d \theta - SC_B \\
&= \int_{-\infty}^{\infty} (208.4 + 459.80\theta - 17.95\theta^2 - 99.84\theta^3) \times \\
& \quad \frac{1}{\sqrt{2\pi} \times 0.0309} e^{-\frac{(\theta - 0.506)^2}{2 \times 0.0309^2}} d \theta - 250 \\
&= 195.82
\end{align*}
\]

and utility of C is:

\[
\begin{align*}
\text{utility of } C &= \int_{-\infty}^{\infty} U_A(r) \pi_C(\theta | x) d \theta - SC_C \\
&= 225.34
\end{align*}
\]

hence (225.34 > 195.82), C is more appropriate than B in the eyes of A.

### 2.2 Open Trust Model

Above discussion is under the condition that the "closed society of agents" must be defined at first, but it is nearly impossible for inherent open and dynamic Web. Our idea is that at every stage of the procedure (after every given observation) one should compare the (posterior) utility of making an immediate decision with the "expected" (preprosterior) utility that will be obtained if more observations are taken. If it is cheaper to stop and make a decision, that is what should be done.

The goal of preposterior analysis is to choose the way of conducting observation (consultant fee) and the cost of conducting observation (consultant fee). Note that these two quantities are in opposition to each other, we propose an approach to balance these two costs.

<table>
<thead>
<tr>
<th>Step</th>
<th>Path</th>
<th>Statement</th>
<th>Reliability Factor</th>
<th>Prior Distribution</th>
<th>Posterior Distribution</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A→D→B</td>
<td>0.55</td>
<td>0.8200</td>
<td>0.6400</td>
<td>0.0800</td>
<td>0.5869</td>
</tr>
<tr>
<td>2-1</td>
<td>A→D→C→B</td>
<td>0.62</td>
<td>0.7200</td>
<td>0.369</td>
<td>0.0492</td>
<td>0.5986</td>
</tr>
<tr>
<td>2-2</td>
<td>A→D→C→B</td>
<td>0.48</td>
<td>0.6480</td>
<td>0.596</td>
<td>0.0390</td>
<td>0.5095</td>
</tr>
<tr>
<td>3</td>
<td>A→D→C→B</td>
<td>0.37</td>
<td>0.4986</td>
<td>0.369</td>
<td>0.0337</td>
<td>0.5095</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step</th>
<th>Path</th>
<th>Statement</th>
<th>Reliability Factor</th>
<th>Prior Distribution</th>
<th>Posterior Distribution</th>
<th>Utility</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A→D→C→C</td>
<td>0.46</td>
<td>0.9000</td>
<td>0.3000</td>
<td>0.1500</td>
<td>0.4676</td>
</tr>
<tr>
<td>2</td>
<td>A→D→J→C→C</td>
<td>0.58</td>
<td>0.8075</td>
<td>0.4767</td>
<td>0.0577</td>
<td>0.5194</td>
</tr>
<tr>
<td>3</td>
<td>A→D→J→K→C→C</td>
<td>0.60</td>
<td>0.6137</td>
<td>0.5194</td>
<td>0.0416</td>
<td>0.5396</td>
</tr>
</tbody>
</table>
A reasonable decision. The expected utility of 219.24 > 216.09, so we prefer to select B instead of C. Hence expected utility of not making immediate decision is

\[
\int_{-\infty}^{0.5914} 207.54 m_c(x) dx + \int_{0.5914}^{\infty} U_A (r) \pi_B (\theta | x) d\theta - SC_B dy - 3 = 219.24
\]

This just is the opportunity cost. Because 219.24 > 217.78, and then further investigation would be worth while money, in other words, A should send request to his acquaintances for their opinions. The residual process of Bayesian sequential analysis are shown in Table 3 and remember that the further exploiting should be terminated immediately along the path on which a cycle is detected. As shown in Table 3, at the stage 3, the expected utility of C becomes larger than that of B, and because 218.06 > 216.09, making an immediate decision is more profitable. Therefore, A should stop investigating and select C as a decision. The advantage of sequential analysis should be clear now. It allows one to gather exactly the correct amount of data needed for a decision of the desired accuracy.

3. ACKNOWLEDGMENT

The work is funded by subprogram of China 973 project (NO. 2003CB317006), National Science Fund for Distinguished Young Scholars of China NSF program (NO. NSFC60533040), and also Program for New Century Excellent Talents in University of Ministry of Education of China (NO. NCET-04-0545).

4. REFERENCES

