Enabling Adaptation in Trust Computations

Longo Luca, Dondio Pierpaolo, Bresciani Riccardo, Butterfield Andrew, Barrett Stephen

School of Computer Science and Statistics, Distributed System Group, Trinity College Dublin, College Green 2, Dublin

Selftrust 2009, November 15-20, 2009 - Athens/Glyfada, Greece
## A Computational Model of Trust

System able to quantify a level of trustworthiness for entities acting in a specific domain. A level of Trust can be used as a filter to select more reliable resources, decreasing the complexity of the environment and increasing the quality of interactions.

## The problem

Due to the not-static property of trust, that changes over time, trust models should consider adaptive strategies to produce accurate and context-aligned degree of agents’ trustworthiness.
Why is Trust Adaptive?

- trustee entities might change their behaviour, producing pieces of evidence that require their level of trustworthiness to be adjusted.
- the external environment might change, producing constraints that might affect the way entities judge other entities’ behaviours (and therefore their Trust levels), or affect the threshold required for collaboration.
The Research Question

*Given a certain distribution of Trust values among a population and given our model of trust, is the trust model able to predict such a distribution?*

*Is it possible to parameterise the trust model, in order to maximise the accuracy of the system’s predictions?*
LTTM - Temporal Trust Model

LTTM considers temporal properties as pieces of Trust evidence to compute the trustworthiness of Wikipedia project’s articles and users.

- activity: agent’s activity compared to the system activity (no. of interactions over time);
- presence: models the human notion of experience over time;
- frequency: interactions compared to agent’s life cycle by using a system constant (interactions expected on a given period of time);
- regularity: models the concept of persistence over time;
Blind-aggregation

The blind-aggregation of the 4 temporal factors is straightforward, but not very accurate.

\[ T(\gamma) = \frac{1}{4} \text{Activity}(\gamma) + \frac{1}{4} \text{Presence}(\gamma) + \frac{1}{4} \text{Frequency}(\gamma) + \frac{1}{4} \text{Regularity}(\gamma) \]

Defeasible reasoning

A more accurate reasoning process takes into account more complex rules by analysing the possible contradictions among basic rules.
Cross tabulation among 4 temporal factors

The set of $4^2$ rules describes all the possible combinations among the factors by assigning the related class of behaviours by considering a boolean value (HIGH (H) or LOW (L)) for each factor (median, mean, std).

The literature in Computational Trust suggests the following five clusters (\textit{TrustClass}):

- Very Trustworthy (VT);
- Trustworthy (T);
- Neutral (N);
- Untrustworthy (U);
- Very Untrustworthy (VU).
Behaviour and rules

A general rule describing a user’s behaviour is a Horn clause:

\[ T(A, P, F, R) \rightarrow TrustClass \]

The rules that impose coherence are of the following kind:

\[ TrustClass_1 \land TrustClass_2 \rightarrow FALSE \]

Validation of inference rules: first order logic theorem prover.

A possible defeasible reasoning cross tabulation

<table>
<thead>
<tr>
<th>AP \ FR</th>
<th>LL</th>
<th>LH</th>
<th>HH</th>
<th>HL</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL</td>
<td>VU</td>
<td>VU</td>
<td>U</td>
<td>VU</td>
</tr>
<tr>
<td>LH</td>
<td>VU</td>
<td>T</td>
<td>T</td>
<td>[U]</td>
</tr>
<tr>
<td>HH</td>
<td>N</td>
<td>VT</td>
<td>VT</td>
<td>T</td>
</tr>
<tr>
<td>HL</td>
<td>U</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Example: the fact that an entity has high experience (P), it shows high frequency (F), but it has a low activity (A) and it does not interact regularly (R), reflects in categorising it as a ‘Untrustworthy’ [U] entity.
Adaptive Trust function

During the formalisation of the Trust function, it is generally unrealistic to assume that all the Trust schemes support the final degree of entities’ trustworthiness equally strongly.

The Trust function need to be self-adaptive and each Trust scheme should have its own weight:

\[ T(\gamma) = W_a \cdot A(\gamma) + W_p \cdot P(\gamma) + W_f \cdot F(\gamma) + W_r \cdot R(\gamma) \]

where \( \gamma \) is a given agent, and \( W_x \) is the weight of the Trust scheme \( X(\gamma) \).
The adaptive process

Trust Schemes
- T1
- T2
- T3
- T4

Reasoning Rules for aggregating Trust Schemes

Test Coherency with Theorem Prover
- yes
- no

Value-class reasoning table

0/1 algorithm

CLASS

TRUST Function
- w1
- w2
- w3
- w4

Normalization / scaling

Matching %
- yes
- no

Confidence threshold

Model Validation

ADAPTIVE algorithm
The problem: finding weights for Trust schemes

Global optimisation problem in a multi-dimensional space.

*Search space:* 4 temporal schemes.

*Supposition:* each temporal factor is bounded in $[0, i]$, our space is $[0, i]^4$.

*Self-adaptive algorithm:* simple heuristic that improves the basic exhaustive search by selecting at each interaction only promising regions of the search space.
Dataset: Finanzaonline the popular Italian finance forum.
- **Online users**: 30,000 users;
- **Online threads**: almost 1,000,000
- **Online messages**: 11,000,000.

Goal: approximate the best choice of the four LTTM’s Trust schemes in order to maximise the accuracy of our model.
Set up

A-priori sampling population distribution: the trustworthiness of each members has been based on an explicit pool promoted in collaboration with the forum administrators.

<table>
<thead>
<tr>
<th>Trust class</th>
<th>Percentile</th>
<th>No. of Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>VT</td>
<td>0.01</td>
<td>313</td>
</tr>
<tr>
<td>T</td>
<td>0.09</td>
<td>2817</td>
</tr>
<tr>
<td>N</td>
<td>0.2</td>
<td>6260</td>
</tr>
<tr>
<td>U</td>
<td>0.2</td>
<td>6260</td>
</tr>
<tr>
<td>VU</td>
<td>0.5</td>
<td>15653</td>
</tr>
</tbody>
</table>

*High/low algorithm*

- threshold for Presence, Frequency and regularity schemes: \( T = average + standard \text{ deviation} \).
- threshold for *activity factor* (distribution results were affected by a very high variance): \( T = average \).
Results: without adaptation

Schemes with 1/4 strength: 45.29% of matching (the sampled population does not seem to behave as we should expect, apart from the very trustworthy entities.)

<table>
<thead>
<tr>
<th>Trust class</th>
<th>Percentile</th>
<th>No. of Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>VT</td>
<td>0.01</td>
<td>313</td>
</tr>
<tr>
<td>T</td>
<td>0.09</td>
<td>2817</td>
</tr>
<tr>
<td>N</td>
<td>0.2</td>
<td>6260</td>
</tr>
<tr>
<td>U</td>
<td>0.2</td>
<td>6260</td>
</tr>
<tr>
<td>VU</td>
<td>0.5</td>
<td>15653</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trust class</th>
<th>Percentile</th>
<th>No. of Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>VT</td>
<td>0.011</td>
<td>342</td>
</tr>
<tr>
<td>T</td>
<td>0.006</td>
<td>216</td>
</tr>
<tr>
<td>N</td>
<td>0.120</td>
<td>3714</td>
</tr>
<tr>
<td>U</td>
<td>0.040</td>
<td>1259</td>
</tr>
<tr>
<td>VU</td>
<td>0.823</td>
<td>25774</td>
</tr>
</tbody>
</table>
We performed experiments with the following metrics:

- \( M \) is the percentage of overall matched entities;
- \( VT \) is the percentage of very trustworthy matched entities (defined in relation to the size of the smallest of the two VT class, i.e. 313 members);
- \( Err \) is the average position error, i.e. the average of the distance \( D \) between the two classes for all the entities in the population: \( Err = \text{avg}(D) \).
Results with adaptation

The 4 parameters that maximised the metric, the optimal, the mean, the worst % of the prediction gained by our adaptation in comparison to the mean.

<table>
<thead>
<tr>
<th>Metric</th>
<th>$W_a$</th>
<th>$W_p$</th>
<th>$W_f$</th>
<th>$W_r$</th>
<th>Optimal</th>
<th>Avg</th>
<th>Worst</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>0.89</td>
<td>0.39</td>
<td>0.11</td>
<td>0.01</td>
<td>60.29</td>
<td>54.44</td>
<td>45.7</td>
<td>10.82</td>
</tr>
<tr>
<td>$VT$</td>
<td>0.20</td>
<td>0.79</td>
<td>0.12</td>
<td>0.11</td>
<td>78.30</td>
<td>45.29</td>
<td>9.58</td>
<td>72.94</td>
</tr>
<tr>
<td>$ERR$</td>
<td>0.89</td>
<td>0.79</td>
<td>0.29</td>
<td>0.09</td>
<td>0.57</td>
<td>0.64</td>
<td>0.93</td>
<td>11.13</td>
</tr>
</tbody>
</table>

- very trustworthy users, with a 78.3% matching;
- overall matching of 60.29%;

Results considering just entities that can actually be matched

Example (VU) Trust classes: minimum value between 25774 and 15653.

Matchable entities for all the classes: 21155. The highest percentage of matched entities (metric $M$, optimal value) was 60.29%, (18875 entities). Since $18875/21155 = 0.893$, almost 90% of the population was correctly matched by the adaptive algorithm.
Good prediction

The preliminary results show a good gain in the quality of prediction, even if a boolean threshold has been adopted to produce the Trust classes in the reasoning-rules based computation.

Promising methodology → further research may be carried on.

Future directions

On-line learning algorithms (strong mathematical background properties).

Multiple regression techniques for comparisons and detailed evaluation among Trust schemes.