Argumentation vs Aggregation of Trust Evidence (Extended Abstract)

Pierpaolo Dondio
Trinity College Dublin
College Green 2
Dublin, Ireland
+353-1-896-2730
Pierpaolo.Dondio@cs.tcd.ie

Stephen Barrett
Trinity College Dublin
College Green 2
Dublin, Ireland
+353-1-896-2730
Stephen.Barrett@cs.tcd.ie

ABSTRACT
The decision to grant trust in virtual societies is often an evidence based process. The evidence for such decision derives from a complex set, where mutual relationships and contradictions are likely, rather than form a list of isolated objects. This paper compares the basic and widely used aggregation strategy, where conflicts and mutual relationships among evidence are ignored, with an argumentation-based strategy.

Categories and Subject Descriptors
I.2.11 [Artificial Intelligence]: Distributed Artificial Intelligence – *intelligent agents, multiagent systems.*

General Terms
Algorithms, Experimentation, Human Factors

Keywords
Computational Trust, Web Communities, Argumentation Theory

1. INTRODUCTION
A *trust computation* is a multi-staged process that produces a level *of trust* for a (digital) trustee agent [1]. This paper focuses on the critical stage of aggregating the various pieces of evidence into a final trust value for the trustee. Since the set of evidence used is diverse and composite, it is likely that the various pieces of evidence, collected using different reasons, may collide, conflict or have some sort of mutual influence, either negative or positive. It seems reasonable therefore that the various pieces of evidence should be combined by taking into consideration their mutual relationships and their logical consistency, in order to reach a more sustainable and correct decision. This means that trust should be the result of an *argumentation* over the various set of evidence rather than a simple *aggregation* of supposedly independent objects. The difference between the two strategies can be substantial, as a single piece of evidence can invalidate a large set, leading to significantly different results as compared to the ones obtained by aggregation.The use of a simple aggregation strategy is widely used and often justified by the limited and well-bound set of evidence that a trust model uses, usually limited to two types of evidence, such as previous interactions history and recommended values. Basic forms of aggregation include global summation, as for example employed in the eBay feedback system [4]. Averaging is another classical aggregation strategy. As a form of linear combination (often weighted), averaging ignores contradictions and blends values. The linear combination recognizes that a piece of evidence has *more importance* than another, but it stills ignores the mutual relationship among them. Moreover, the produced value hides different opinions. A more sophisticated aggregation strategy considers the standard deviation (or more advanced indicators of variance) of the values of the various pieces of evidence, in order to have a quantification of the consensus among them, stored as an uncertainty value, such as in [2]. However, still no relationships among evidence are directly considered.

Cite as: Aggregation vs. Argumentation of Trust Evidence. *Pierpaolo Dondio, Stephen Barrett*, *Proc. of 9th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, van der Hoek, Kaminka, Lespérance, Luck and Sen (eds.), May, 10–14, 2010, Toronto, Canada, pp. 1567-1568

2. SETTING UP THE PROBLEM
We assume that a level of trust is computed using various pieces of evidence that form arguments to sustain the trustworthiness of the trustee. For instance, a trust models can assess the trustworthiness considering the following 3 arguments, corresponding evidence: past-outcomes, reputation and longevity in the domain. general shape of the model is depicted in figure 1 and 2. Figure 1 illustrates the aggregation scenario, where each argument is isolated, while figure 2 shows the argumentation scenario, with its potential attack and support relationships.

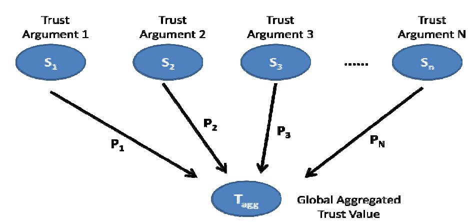

Figure 1 Aggregation strategy

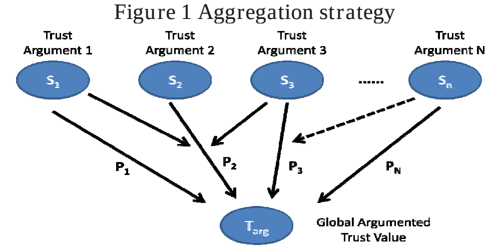

Figure 2 Argumentation strategy

Each argument is assumed to have a strength s in [-1,1]. A positive value suggests that the argument is in favor of the trustee and vice-versa. Each argument has an associated second value, the value of plausibility, which quantifies the validity of each argument in the context. Alice can have a very high strength for *reputation*, but the use of *reputation* could be a totally invalid trust argument since, for instance, it is known in the specific context that reputation values are strongly biased. The plausibility value thus acts as a weight for the strength of the corresponding argument. The two strategies we seek to compare are defined as follows. The first one, aggregation, considers each trust argument independent. The plausibility values p_i are used to weight each argument's strength only. The global *aggregated* trust value for a trustee is given by a weighted average:

$$T_{agg} = \frac{1}{n}\sum_{i=1}^{n} p_i s_i \qquad (1)$$

In the argumentation approach, arguments can attack or support other arguments. Note how it is not the argument that is attacked directly, but its plausibility. The result of the attack/support is the modification of the value of plausibility p_i, potentially nullifying it if the attack comes from a strong enough argument.

The argumentation strategy is computed in the following way. First, a defeasible semantic (described in the next section) is applied according to the mutual relationships identified among the arguments. The final *argumented* trust value is again computed using formula 2, but now with modified values of p_i.

3. SEMANTIC AND RELATIONSHPS

A semantic requires the definition of the *function of conclusion* f_c, the *function of attack* and *support* and the *accrual* of reasons, that defines how to compute the strength of a conclusion based on *n* independent arguments of strength S_n.

Function of Conclusion

Given a premise *A* of strength S_a and plausibility P_{ac}, the strength of the conclusion (see [3] for justification) is:

$$S_c = S_a P_{ac} \qquad (2)$$

Function of Attack and *Support*

These two functions define how the strength of a reasoning link P_{ac} changes if attacked by an argument of plausibility P_b and strength S_b . Note again how S_a is not involved in the process. We used the following function of support:

$$S_b P_b > S_a P_{ac} (activation\ rule) \qquad (3a)$$

$$P_{ac} = P_{ac} + c_1 (P_b - P_{ac}) S_b \qquad (3b)$$

With analogous argument, the function of attack is the following (with no activation rule):

$$P_{ac} = P_{ac} - c_2 P_b S_b \qquad (4)$$

Accrual of Reasons

In our simple inference network (figure 2 and 3), the only accrual of reasons is actually how to compute the final trust value, represented by formula (1) of previous section.

Mutual relationships

A mutual relationship defines if an argument attacks/supports another. It has the form Arg1 attacks/support arg2. Each argument is required to have the proper positive or negative strength, and each relationship has a plausibility value attached. Let's analyze some of the identified rules

$$AR_1: Activity^+ \rightarrow Longevity^+, 0.9$$

$$AR_2: Activity^- \nrightarrow Longevity^+, 0.9$$

AR_1 states that *longevity* is a stronger argument when the trustee is also active, while a low degree of *activity* invalidates a high degree of *longevity* (AR_2). With the same notation we identified also:

$$AR_3: Activity^+ Stability^+, 0.9$$

$$AR_4: Activity^- \nrightarrow Stability^+, 0.9$$

Stability is stronger evidence if the entity is *active* (AR_3), while even a high degree of *stability* is undermined if the entity is not *active*, since it could be *out of business* rather than complete.

A lack of pluralism weakens high activity, since it undermines its validity and controllability:

$$AR_5: Pluralism^- \nrightarrow Activity^+, 0.75$$

A lack of longevity attacks also a high level of activity. It is a common used metric that spammers are likely to create a large amount of activity for short time and suddenly disappear:

$$AR_6: Longevity^- \nrightarrow Activity^+, 0.9$$

4. EVALUATION

We compared an argumentation and an aggregation approach on top of the trust computation we performed over 7718 Wikipedia articles described in [5]. We quantified the effect of the two strategies by comparing each of them with the internal Wikipedia quality system of featured and standard article. A strategy performs well if it gives high score to featured articles and average score to standard articles. The metrics used for quantifying each strategy are the percentage *%FA* of featured articles predicted correctly, the percentage %SA of standard article predicted correctly, the correlation *C* between the distribution of standard articles and featured articles according to their trust value. If C is low, featured and standard articles have been efficiently divided. Inputs of our evaluation are the strength and plausibility of a set of trust argument computed for each article. The arguments are the following (see [5] for details): *Stability, activity, pluralism, Similarity, Standard-Compliance* and *Authorship.*

Table 1 Argumenation vs. Aggregation

	C	*%FA*	*%SA*
Aggregation	21	77.9	16.4
Argumentation	18.4	80.8	13

As the tables illustrate, the argumentation strategy outperforms the simple aggregation, augmenting the quality of the computed trust values. The correlation decreases from 21% to 18.4%, meaning that the system is more efficiently dividing featured and standard articles. The percentage of featured articles that are now recognised trustworthy is increased by 2.9%, (3.6% relatively) while the percentage of standard articles not considered trustworthy is decreased by 3.4% (20.7 relatively). The gain of performance is due mainly to the following 2 argumentation rules. First, low activity that invalidates stability, useful to spot stable articles that are abandoned rather than mature and complete (AR_4, applied over 547 articles). Second, lack of pluralism that invalidates high activity (AR_6, applied over 298 articles), useful to spot articles created by few people, usually out of control

CONCLUSIONS

In this paper we evaluated the impact of an argumentation process as a technique to better combine contrasting sets of trust evidence. The results obtained show an increase of the quality of the trust prediction from 3% to 20%.

References

[1] D. Gambetta. Can we trust trust? Making and Breaking Cooperative Relations, pp. 213–237, 2000.

[2] J. Sabater, C. Sierra. REGRET, A reputation model for gregarious societies, 4th Workshop on Fraud and Trust in Agent Societies, Montreal, Canada. pp. 61-69, 2001

[3] J. Pollock. Defeasible Reasoning with Variable Degrees of Justification. Artificial Intell., N. 133, pages 233-282, 2001

[4] Ebay.com, Online Auction website

[5] Pierpaolo Dondio. Trust as a form of Defeasible Reasoning, Phd Thesis, Trinity College Dublin