Paper

On Subjective Trust Management

Michał Majdan

Abstract—Trust and reputation management is gaining nowadays more attention then ever as online commodity exchange and other open virtual societies became a widespread reality. Most widely used computational models use reputation metrics as global property assigned to each party. More sophisticated models try to use reputation as subjective property. While introducing subjective reputation there arise a need to model preferences of agents. In this paper we propose to use weighted ordered weighted average (WOWA) operator to support the decision maker in assessing available evidence about other's party behavior. The WOWA aggregation is defined by two weighting vectors: the preferential weights assigned to the ordered quantities and the importance weights assigned to several attributes. It allows one to express both the preference regarding sources of information by the corresponding importance weights and the compensation between attribute values aggregated by the preferential weights.

Keywords*— *decision support, management concepts, multicriteria reputation model.

1. Introduction

Internet society has become reality. Nowadays more and more of commodities are exchanged via online commerce sites. Success of online auctions like e-Bay or Allegro has shown how dramatically fast more and more aspects of human life are moving online. But commerce of goods and services is not the only domain changed by the Internet. Peer-to-peer (p2p) networks exchanging software and data are getting more and more attention too. Wikipedia, an online encyclopedia constantly developed by anonymous users sharing their wisdom end expertise with others. We rely more and more on people or agents whom we probably are never going to meet face to face in "real" life, probably we are not going to establish long term relationship with most of them. Therefore old ways of assessing transaction counterparts are of little use in online world. Thus users try to assure security for online life by designing and implementing trust management systems, algorithms and protocols that are to provide trust between members of such systems.

There are various definitions of trust and there are differences among researchers on how define trust. Most popular in literature is definition by D. Gambetta [1]:

"*Trust is a particular level of subjective probalility with which an agent will perform a particular action both before he can monitor such action (or independently of his capacity of ever to be able to monitor it) and in context in which it affects own action*".

The "probability" mentioned above is "subjective" since according to this definition the same agent may be perceived differently by other members of the same virtual community. Not all models follow this assumption some aim to compute some kind of global evaluation of a certain agent (it will be mentioned below).

The second component that is very tightly coupled to the concept of trust is reputation. Generally reputation is "what-is-said" about certain agent in a given community. That's why if we consider any subset of this community the reputation of the same agent may differ. It's important to notice that reputation is not what a certain agent "thinks" about the other, reputation is a kind of aggregation of opinions communicated by a given group of agents. This concept coming from the cognitive theory of reputation is exploited by J. Sabater *et al.* in [2].

2. Trust and Reputation Models

Computational trust and reputation management models usually use 4 sources of information to build view on a particular agent:

1. Own direct experience – these are valuations that come from agent's own interactions with the target agent. The credibility of those valuations is perfect, so they are most valuable for any trust model. The problem is that in real world situations rarely there exists more than 1 interaction that involves the same 2 agents. So the ability to collect enough data based on direct experience is very limited.

2. Observations – in some environments agents are able to observe performance of the others. Like with the direct experience those evaluations are highly valuable. Very rarely though open virtual communities offer the ability to monitor interactions of others to everybody.

3. Witness information – these are valuations provided by other agents regarding third party. Those can be vulnerable to untrue valuations malicious agents who are willing to destroy others reputations or contrary inadequately promote others. Apart from those threats different agents assess others performance differently. Behavior totaly inappropriate in the eyes of one can be correct when assessed by the other. Witness information is actually something that constitutes reputation as word-of-mouth.

4. Context information – each interaction is performed in a particular context. Context can be described as a vector of parameters different for each system. For online auction example this can be the value of the transaction, whereabouts of the seller or buyer, means of payment, type of merchandise.

Trust models usually recognize at least 2 roles of an agent that could be considered when determining trust metrics. These are firstly trust on whether particular agent will perform as we expect him to do and secondly whether he can be trusted as an informant. This can be further divided into roles based on what kind of activity we require from the particular agent: for example, one can be trusted seller but fail at delivering services. Other important dimension that is used in trust and reputation models is time. Usually trust is a function of reputation and the time elapsed from the events that update the reputation. This reflects belief that conclusions based on past behavior are less accurate the further in the past we look. What differs among various reputation based models is what kind of so called third party information is exchanged.

Some of the models assume that peers exchange their overall evaluation of "other party" some call it image [2]. In such situation whole evaluation process as well as the criteria may be totally different at each peer. But clear advantage of such models is that the information traffic is minimal.

Second kind of reputation based models assume that agents exchange evaluations on transactions they had with the other party like in e-Bay feedback system. This allows to incorporate time into the model as receiver of such information is able to for example disregard evaluations associated with transactions older then arbitrary threshold. Such models are most widely used in online commerce sites, for example e-Bay, we will look into this later. Still such models assume that criteria each evaluator uses and his utility is not known and not shared among members of such virtual community.

Another thing that needs to be considered while designing trust model is source reliability. As it was stated above, most of the information that particular agents has to rely on is information provided by third party. It's crucial to eliminate information coming from unreliable sources. The source may be lying on purpose in order to increase own reputation or reputation of some associated identity or to decrease others party reputation. Models developed so far address this problem various ways. One way is by introducing pre-trusted peers [3] and forcing agents to put at least some amount of "trust" in the pre-trusted peers. Others set up protocol to detect malicious peers by asking them about things already known to the asking party. If answer is different such peer's reliability as an informant is decreased. This approach is used in models that assume that the same transaction may be observed by more then 2 involved parties.

One other thing that can be considered in trust/reputation models is agent's confidence about his judgment. When asked about opinion on some other party agent can provide his rating noting that he is sure about this opinion only up to the certain level. This means that each rating is a vector of two values. This is especially applicable to models using reputation as a message exchanged between parties. In such case agent can express using confidence the fact that he had very little experience with the assessed party. Models using transaction rating do not require this additional information. Confidence value associated with specific valuation make the aggregation procedure significantly more complex. Up to now quite large number of trust models have been developed. Most important are presented below. We have focused on how user preferences are incorporated in those models and how valuations are aggregated.

3. Selected Trust Management Concepts Review

3.1. Online Auctions Reputation Models

The most widely used are very simple reputation based systems like e-Bay feedback system. They are based on three valued feedback provided by transaction parties on how they assess the transaction. Positive values mean that the party is satisfied with partner's performance. Neutral values usually mean that transaction is judged as sufficiently correct. Negative feedback values usually mean that transaction was highly unacceptable. It turns out that vast majority of feedback is positive with very little neutral or negative feedback. But there exists significant percent of transactions with no feedback at all. The problem is that giving negative or neutral feedback threatens that other party will retaliate, so people who actually dislike the performance of their transaction partner choose not to provide feedback instead of giving negative. Transaction evaluations (–1,0,1) are summarized over certain period of time (e-Bay – 6 months). Although simple and vulnerable to false information the above mechanism proved to be usable since online auction and general open e-Commerce sites are successful.

Preference structure representation. User preferences are not expressed directly but through feedback value. It's not known what made agent to provide certain value.

3.2. Probabilistic Models

The eigen-trust [3] algorithm has been proposed as a mean to provide trust in the anonymous peer-to-peer networks. The measure of trust is a normalized value between 0 an 1 reflecting number of satisfactory and unsatisfactory transactions with a given peer versus all peers. Since not all peers interact with each other the experiences of others are the only source of information. The party gathering information may ask his direct counterparts about their opinion on a specific target. Their opinions are weighted by the trust value that given peer places in them. Doing further this way one party may ask his friends about their opinions. This is done by multiplying consecutive local trust values. Such computations converge under certain assumptions to the global trust vector that shows how much trust the system as a while places in its members.

The eigen-trust model uses concept of transitive trust. Valuations exchanged between peers refer to the peers reputation not individual transactions. It is assumed that the preference structure of each peer is the same, namely it's the difference between successful interactions and unsuccessful ones. The only piece of subjectivity is when agent is to evaluate whether certain interaction was successful or not. Another example of probabilistic approaches is model proposed by Shillo *et al.* [4]. In this model each interaction is assessed as either good or bad. The personal trust value that one agent places into another is $Q = e/n$, where n is a number of observed interactions and e is a number of positive ones (when other party behaved honestly). An agent can ask other agents about their impressions on observed interactions. Model assumes that agents do not lie about interactions they observe therefore while considering messages from others, there is no need to resolve conflicts (the same interactions assessed differently by different agents).

3.3. Abdul-Rachman, Heils Model

Some models use linguistic labels instead of numerical ratings to represent social valuations. This model is especially interesting as it tries to deal with the issue of different perception of the same information by different agents (people). The idea of this model is that assessing agent can give the other party one of 4 labels based on his opinion about the other. These labels are namely:

– very trustworthy (VT),
– trustworthy (T),
– untrustworthy (U),
– very untrustworthy (VU).

Agent maintains number of interactions of each of the above categories for every other agent in the society (system). The general assessment of the agent is label with maximum number of interactions that "support" this label. In case of two or more labels having the same count model introduces uncertaintly measures into assessment.

The most interesting part of the model specification is how it deals with the reports (opinions) provided by other members of the system. If an agent receives information from the other agent assessing some third party as for example VT and the agent has previously assessed the same other party as only T he will adjust any future opinions received from the same agent accordingly, that is he will lower them. The preference structure at this model is not externalized by the agents but it is expressed indirectly by comparing assessments of the same information.

3.4. OWA Trust

In 1988 Yager [5] proposed new aggregation operator ordered weighted average (OWA). It is similar to weighted mean operator. Weights though are not assigned to particular criteria but to the values within criteria permutation from the biggest value to the smallest value. Formal definition of the operator is as follows: given vector of n criteria we have x_i for $i = 1 \ldots n$ and preferential weights vector w_i for $i = 1 \ldots n$ while $\sum_{i=1}^{n} w_i = 1$. The OWA operator is defined as follows:

$$\sum_{i=1}^{n} w_i x_{\sigma(i)} ,$$

where $\sigma(i)$ is certain permutation of vector x that

$$x_{\sigma(1)} \geq x_{\sigma(2)} \geq x_{\sigma(3)} \geq \ldots \geq x_{\sigma(n)} .$$

This operator is used in a trust management model proposed in [6]. Local reputation values are calculated as $t_{ij} = 1$ or $t_{ij} = 0$ if the given peer would have only one successful or unsuccessful interaction with the other peer. If there was more interaction between two peers the overall local reputation is calculated as follows:

$$r_{ij}^{n} = \alpha^{n} r_{ij}^{n-1} + (1 - \alpha^{n}) t_{ij}^{n} .$$

The α^n parameter is accounts for aggregation "freshness", that is how past transactions are important compared with the last one. High values are for the case when past values are far more important than the last one otherwise last interaction is more important. Local reputation is of little use if interactions between same peer pairs are rear. This is almost always the case in online society systems. For such situation model provides the concept of network reputation. Local reputation values from the pool of voting peers are aggregated using the OWA technique. Local reputation values are ordered and aggregated using the set of weights. Authors do not impose any procedure on selecting weights for a given model realization. Authors show that selection of weights reflects decision maker's attitude towards situation.

4. Multicriteria Reputation Model

The model presented in this paper is addressing the potential problem associated with reputation representation. As it was mentioned above models exchange either information about other agents or about transactions. In both cases they exchange aggregated social evaluations build on how they personally perceive other users or certain transactions. Such approach hide the preference structures of the agents. Messages biased by the ones preference structure may lead to wrong assessment of other agents. The proposition is to make information exchanged between agents as objective as possible and to allow consumers of this information to make their own judgments. Models that assume exchange of social valuations of other agents bias the evaluations more than ones exchanging data on transactions so in the presented concept the content of the message will be associated with the single transaction.

4.1. Outline

The first thing to do implementing the model is to set up a set of criteria associated with the single transaction.

Next we need a common way to express the preference structure of each member of the system. This need to be as easy as possible and selected a priori since in a general case agents may not be able to modify their preference structures assessing each interaction.

We need to set up an algorithm to aggregate interactions and make a decision based on the criteria whether to trust or not to trust.

4.2. Selecting Criteria

Setting up a group of criteria is the most vulnerable part of the model. It is very tightly coupled with the actual domain the model being implemented into. The important part is that calculating satisfaction levels for this criteria should be as objective as possible. If it is not we are threatened to exchange single biased social valuation into a set of them which is far less attractive but still can provide some advantage. We can consider following example criteria:

- latency of payment or delivery in case of an auction service;
- price (compared to other items of the same kind in the service in the same time);
- number of e-mail's exchanged between buyer and seller;
- response time in case of a question;
- download speed in case of p2p network;
- number of errors during download, etc.

For successful application of aggregation operator, criteria satisfaction measures need to be commeasurable, therefore they need to be normalized. Since selection of criteria is out of the scope of this paper this topic won't be elaborated further.

4.3. Setting up Preference Structure

The preference structure implemented in the presented model is twofold. First setup is associated with the issue of aggregating information about past transactions reported by other users. This is actually aggregating reputation metric with regard to each of the criteria separately. Every trust management model based on reputation has its own way of aggregating reputation data. This aspect is discussed in detail in [7]. Previous section gives short overview of this topic with regard to the selected models described above. In this paper author proposes one possible aggregation technique with very significant expression power.

An important aspect to consider is source reliability. For now we do not go far into this area. As it was mentioned in the beginning there are being developed various algorithms and protocols that deal with this problem. For now we assume that given agent is able to assess the reliability of the information coming from other agents, and eliminates messages from malicious informants. It's worth to notice that being a malicious informant may not come with failing with delivering as expected. Presented model is to deal with assessing trust with regard to the agent performance in delivering merchandise, data, payments ... and not reputation valuations (criteria satisfaction levels) associated with other members of the community.

The next aspect addressed by trust management models is information being biased by informants own attitude towards assessed situations. This is the problem that presented model is trying to answer. As mentioned above satisfaction levels of the selected criteria should be objectively calculated possibly by the system itself, or normalized in order to establish common view of this values by all agents. The next problem is how agent may express his preferences in order to judge whether he can trust the other or not. This involves comparing calculated trust value with arbitrary threshold (independent for each agent) in order to make a decision to go on with the interaction or to retreat. Setting up preference structure should be as easy and understandable but at the same time needs to have as much expressive power as possible to cover possibly wide range of preferences.

4.4. Weighted Ordered Weighted Average

Weighted ordered weighted average (WOWA) aggregation operator was proposed by V. Torra [8] as a generalization to previous ordered weighted operator proposed by Yager [5]. Torras conception is based on two weighting vectors:

- preferential weights vector w_i[1], associated with values permutation from the highest value to the lowest;
- importance weights vector p_i associated with each of the aggregated criteria.

Formal WOWA definition is described below:

$$WOWA(x_1, \dots x_n) = \sum_{i=1}^{n} \omega_i x_i,$$

while weights ω are constructed as follows:

$$\omega_i = w^*(\sum_{j \le i} p_{\sigma(j)}) - w^*(\sum_{j < i} p_{\sigma(j)}).$$

Function w^* interpolates points $(i/n, \sum_{j \le i} w_j)$ and point (0,0) if points can be interpolated using linear functions they should be interpolated in this way. In case when preferential weights p_i are equal and sum up to one, WOWA becomes standard OWA operator with preferential weights w_i. When preferential weights are equal and sum up to one, WOWA operator becomes weighted average operator. The WOWA aggregation generalizes both OWA and weights average and its actual value are always somewhere in between those aggregation methods.

4.5. Aggregation of Interactions

We need to consider how information regarding past interaction involving assessed agent is being aggregated. We

[1] In general case number of preferential weights can be higher than the number of values aggregated [9].

should aggregate values for each criterion to present to user single vector of aggregated criteria satisfaction levels. We propose to consider use of the WOWA aggregation technique as well suited approach for this case. As mentioned above particular agent while assessing reliability of the informer can use several techniques. However the problem itself is not an easy task and very often agents cannot a priori assess the reliability of others, especially in environments when very rarely there exists more than one interaction involving particular pair of agents. If environment/system allows to monitor reliability of the information, agent can derive a kind of rating for messages he acquired from other parties. If no such functionality is available, messages can be arranged with respect to the date they were created.

In either case derived ratings can be used to establish the vector of importance weights as described in WOWA definition. WOWA gives us also possibility to express attitude towards values of selected criteria satisfaction levels. If particular agent requires that all messages regarding interactions with other party should show high level of performance with respect of a given criteria, he should set preferential weights to form an “anding” operator. If he requires at least one message to be highly satisfactory he should form more “oring” operator. The question of “orness” and “andness” are described in [5].

4.6. Aggregation of Criteria

At this point decision maker is presented with a set of aggregated criteria satisfaction levels. This actually sets up a multicriteria analysis situation. If the situation would be to choose counterpart that can be most trusted with regard to this situation the whole multicriteria analysis methodology and tools could be used to support decision maker (DM) at choosing the best alternative. For instance in case of human DM the reference point method could be successfully applied. But in most trust management situations there exists only a pair of agents and for each of them the interaction requires binary go/not go decision. Further in many cases agent (DM) is not human and cannot modify its preferences interactively. Rather it can be equipped with some a priori set preference structure to filter the available data and make decisions comparing calculated metrics with arbitrary selected threshold. Again WOWA can be used as a scalarizing function. Agent, for example user of online auction service, can establish importance weights to express tradeoffs between criteria. The preferential weights in this are case also used to express how many criteria in general have to be satisfied and to what extent. Calculated overall value is then compared against individual threshold and decision to go with the interaction is based upon this.

5. Conclusions and Further Work

This paper presents outline of trust management model that puts stress on subjective interpretation of information in decision making process. It shows possible way of setting up preference structure by parameters of WOWA aggregation operator. Model requires that criteria selected to describe single interaction can be objectively measured and subjectively assessed. While constructing vector of criteria satisfaction levels it has to be assured that its components are commeasurable.

There are still issues that have to be elaborated. Objective selection and measurement of criteria much depends on particular system and environment but is crucial for successful application of above technique. Some work needs to be done to check real life scenarios if such criteria can be monitored with satisfactory level of objectivity. In case of interaction aggregation setting up importance weights is much dependent on agents reliability verification method. If such method is nonexistent or can be compromised easily it might be good to consider setting all importance weights to equal value reducing WOWA to OWA operator.

It’s planned to apply this model to a real life situation of online auction service. If all agents are forced to express their preferences in a common and understandable form there arises opportunity to analyze users of such system with regard to their preference structure and to adopt system accordingly.

Acknowledgement

This paper has been partially founded by the Polish Ministry of Science and Higher Education under the research grant N N516 4307 33.

References

[1] D. Gambetta, “Can we trust trust”, in *Trust: Making and Breaking Cooperative Relations*. Oxford: Basil Blackwell, 1988, pp. 213–237.

[2] I. Pinyol and J. Sabater-Mir, “Arguing about reputation the lrep language”, in *Eighth Ann. Int. Worksh. Eng. Soc. Agents World ESAW’07*, Athens, Greece, 2007, pp. 192–207.

[3] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, “The eigentrust algorithm for reputation management in p2p networks”, in *Proceedings of the 12th International Conference on World Wide Web WWW’03*. New York: ACM, 2003, pp. 640–651.

[4] M. Schillo, M. Rovatsos, and P. Funk, “Using trust for detecting deceitful agents in artificial societites”, *Appl. Artif. Intel. J.*, vol. 14, no. 8, pp. 825–848, 2000.

[5] R. R. Yager, “On ordered weighted averaging aggregation operators in multicriteria decision making”, *IEEE Trans. Syst. Man Cybern.*, vol. 18, no. 1, pp. 183–190, 1988.

[6] R. Aringhieri, E. Damiani, S. De Capitani Di Vimercati, S. Paraboschi, and P. Samarati, “Fuzzy techniques for trust and reputation management in anonymous peer-to-peer systems”, *J. Amer. Soc. Inform. Sci. Technol.*, vol. 57, pp. 528–537, 2006.

[7] J. Sabater-Mir and M. Paolucci, “On representation and aggregation of social evaluations in computational trust and reputation models”, *Int. J. Appr. Reas.*, vol. 46, no. 3, pp. 458–483, 2007.

[8] V. Torra, “The weighted OWA operator”, *Int. J. Intel. Syst.*, vol. 12, pp. 153–166, 1997.

[9] W. Ogryczak and T. Śliwiński, “On optimization of the importance weighted OWA aggregation of multiple criteria”, in *International Conference Computational Science and Its Applications ICCSA 2007. LNCS*, vol. 4705. Heidelberg: Springer, 2007, pp. 804–817.

Michał Majdan received the M.Sc. degree in computer science from the Warsaw University of Technology, Poland, in 2003. Currently he prepares his Ph.D. thesis in computer science from the Institute of Control and Computation Engineering at the Warsaw University of Technology. He is employed by the National Institute of Telecommunications in Warsaw. He has taken part in projects related to data warehousing and analysis for a telecommunication operator. His research focuses on modeling, decision support, trust and reputation management.

e-mail: Michal.Majdan@elka.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

e-mail: M.Majdan@itl.waw.pl
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland