Decision Support Systems 54 (2013) 929–940

Contents lists available at SciVerse ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Improving computational trust representation based on Internet auction traces ☆

Adam Wierzbicki [a,*], Tomasz Kaszuba [a], Radoslaw Nielek [a], Paulina Adamska [a], Anwitaman Datta [b]

[a] Polish-Japanese Institute of Information Technology, Warsaw, Poland
[b] Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Received 18 May 2011
Received in revised form 8 June 2012
Accepted 23 September 2012
Available online 28 September 2012

Keywords:
Trust management
Reputation system
Text mining
Natural language processing
Sentiment analysis
Classification
Taxonomy
Reference Point methodology
Detailed Seller Rating

ABSTRACT

Computational trust representations are used by Trust Management (TM) systems to elicit information from users about the behavior of others. In most practically used TM systems, simple computational trust representations dominate, such as the three-valued discrete scale of "negative", "neutral" and "positive" used in reputation systems of Internet auctions. This paper asks the question: what is the appropriate system for computational representation of human trust? In order to find an answer, we study a large trace of feedbacks and textual comments from a reputation system of an Internet auction. We discover that users systematically try to add information in the textual comments. Text-mining and NLP approaches reveal a taxonomy of non-positive feedbacks and an importance order on the categories of non-positive behavior. This importance order is further supported by survey data. Based on these observations, we propose and evaluate a complete, new computational trust representation system inspired by the work of Yager. This system is complemented by operators that can be used to produce rankings of most trusted agents. The operator used to create rankings selects Pareto-optimal agents with respect to the multiple criteria revealed by our trace analysis. The proposed system takes into account all criteria utilized by auction users to evaluate behavior, and the relative importance of these criteria. The proposed system is compared to the Detailed Seller Rating system introduced by eBay.

1. Introduction

Trust is crucial for many social and commercial interactions. In online and virtual world settings, one may often interact with other unfamiliar individuals or entities. A comprehensive Trust Management (TM) system can facilitate decision support in such situations.

Most current TM systems use simple computational representations of trust. For instance, many internet auction sites use a three-valued discrete scale of "negative", "neutral" and "positive" (with the notable exception of the new system used by e-Bay). The epinions recommendation system [12] likewise uses a three-valued scale. The FilmTrust [5] recommendation system uses a 10-valued discrete scale. Other theoretical reputation systems often use a continuous scale from [0,1] or [−1,1] [11,16], while a more complex system has been proposed by Josang [9] which uses a two-dimensional scale with uncertainty representation. However, even Josang's system only uses a simple scale for the actual representation of trust expressed by users. In practice, there are multiple aspects, some (perceived) more important than others, which cumulatively and subjectively determine one's trustworthiness. For instance, in an online auctioning scenario, whether the product being sold is genuine (w.r.to what was advertised) is important, but so may be other issues, for instance: whether it was delivered on time and undamaged, and if not, whether the seller was seen to be responsive in amending the situation or not, etc.

Several things are apparent from this illustrative example. Firstly, a simple scale is inadequate to capture the multi-faceted nature of interactions, which cumulatively determine the trustworthiness. Secondly, one can potentially come up with any arbitrary list of criteria, and moreover, different application domains will have different set of criteria. One size does not fit all, and an arbitrary multi-dimensional scale will also be inadequate. Ultimately, the set of criteria should be application specific, and driven by data (users' perceived needs and priorities).

To that end, the questions motivating this paper are: *what is an appropriate system for computational representation of human trust (in the context of internet auctions)? Are the simpler systems used so-far (in)adequate for this purpose? And if they are not, can we propose a better computational representation and ways to process it in a Trust Management system for Internet auctions?*

In our attempt to answer the above posed questions, it is necessary to use information on *how real users evaluate trust or distrust*. One important challenge here is that most available datasets are obtained from TM systems that use a simpler computational trust representation. Nonetheless, these systems, probably in order to account for the limitations introduced by the simplistic computational trust representation, often allow users also to add an accompanying textual comment. The crux of our approach is to dig into such textual comments from data obtained from a real auction site to identify the various issues that the users perceive are important. Specifically, we will investigate the following questions: *do users systematically try to add more information than is allowed by the*

☆ This research has been supported by the Polish Ministry of Science grant 69/N-SINGAPUR/2007/0.

* Corresponding author.
E-mail address: adamw@pjwstk.edu.pl (A. Wierzbicki).

0167-9236/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.dss.2012.09.016

computational trust scale used by the reputation system? *What kind of information is being added by users*? *How are the quantitative reports* (*using the current computational trust scale*) *related to the additional information in textual comments*? In order to answer these and like questions, we employ Natural Language and Sentiment analysis of the textual comments that extracts their emotional content, as well as a classification of comments that reveals an implicit, more complex preference structure of behavior valuations. This preference structure forms a partial order, because it is a dominance structure established by the simultaneous use of many diverse criteria to evaluate behavior.

Based on the analysis of the dataset, we will attempt to propose a trust representation system that is adapted to the implicit user requirements revealed by our analysis. The proposed system is inspired by the work of Yager [26] and Sabater-Mir [21]. It is also in some respects similar to eBay's new Detailed Seller Rating system.[1] Therefore, we will review the DSR system and discuss some of its deficiencies. We will also introduce operators that allow us to process the new computational trust. In this aspect, we go beyond the work of Sabater-Mir, introducing new operators and demonstrating how the proposed computational trust can be used to create rankings of sellers or buyers in Internet auctions, which is crucial for decision support.

The most important discovery from the analysis of our dataset is that auction site users utilize multiple criteria simultaneously to evaluate behavior of others, which in turn leads to a new requirement for the selection operator used to create rankings of buyers or sellers. This operator should choose agents that are Pareto-optimal with respect to the multiple evaluation criteria. We define such an operator based on the Reference Point methodology and compare the rankings it produces to DSR rankings.

Our new computational trust representation, designed for Internet auctions, can be generalized for other applications based on the observation that our proposed criteria are related to norms of behavior. The operator used to create rankings selects agents that are most compliant with multiple norms of behavior, i.e., Pareto-optimal with respect to procedural fairness. Even the importance order of our criteria can be expressed in norms using modal logic.

The rest of this paper is organized as follows: in the next section, we discuss in more detail the research hypotheses of the paper and their contribution to the research questions outlined in the introduction. In Section 3, we introduce the Internet auction dataset and present the relevant results of its analysis that demonstrate that users systematically try to add new information to the system. This information is present in the non-positive feedback and can be categorized into a taxonomy that can be fully expressed using the proposed new computational trust representation. Categories of feedbacks are related to norms of behavior that are the basis of criteria for behavior evaluation. In Section 4, we validate hypothesis 6 by showing that the data supports a second conclusion about the new trust representation: the various categories of non-positive behavior are not equally important to users. Users distinguish that some norms of behavior are more important than others. This information should be reflected in the way the reputation system processes the new proofs. Therefore, in Section 5 we introduce a new computational trust representation that generalizes eBay's DSR system, and new operators that can process the new proofs in a way that takes into account the importance of applied evaluation criteria. In Section 6, we evaluate the new proposed operators using simple simulations that compare them against the ranking used by eBay's DSR system. This evaluation validates hypothesis 7 by showing that it is possible to create a seller ranking based on the proposed proofs in a way that takes into account the importance of behavior evaluation criteria. In Section 7, we present related work on computational trust representation. Section 8 concludes the paper and discusses future work.

[1] http://pages.ebay.co.uk/services/buyandsell/powerseller/criteria.html.

2. Hypotheses about usage of reputation systems in internet auctions

In this paper we will investigate the following hypotheses:

1. Users systematically add significant information in non-positive feedbacks in reputation systems for Internet auctions.
2. The non-positive feedbacks can be categorized into a taxonomy.
3. Each of the leaf categories specifies a criterion for behavior evaluation.
4. Each of the leaf categories is related to a norm of behavior.
5. The criteria of behavior evaluation specified by non-positive feedback categories can be used independently.
6. The criteria for behavior evaluation specified by non-positive feedback categories are not equally important to users.
7. It is possible design a reputation system so that it takes into account more criteria of behavior evaluation and ranks sellers taking into account criteria importance.

In the rest of this paper, the hypotheses listed here will be investigated sequentially, as the validation of the previous hypotheses is usually a condition for the consideration of the next one.

Hypothesis 1 can be verified by an investigation of a large set of feedbacks from a reputation system for an Internet auction. If it can be found that the non-positive feedbacks contain meaningful information that enhances or even changes the interpretation of the feedback value, the hypothesis will be positively validated. This hypothesis is a basic step to consider whether the enhancement of the reputation system is actually required.

Hypothesis 2 can be verified if most non-positive feedbacks from a large set can be assigned to a category of feedbacks, and if the categories form a taxonomy of more general non-positive categories. The leaf categories of the taxonomy should be associated with classification rules (for example, based on regular expressions) that can be used to assign each non-positive feedback into a category.

This hypothesis concerns the organization of the new proposed reputation system. If the non-positive feedbacks would be largely unrelated, it would not be possible to help users to report behavior in a systematic manner. It would also be very hard to predict the content of the future non-positive feedbacks. On the other hand, the existence of a taxonomy of non-positive feedbacks suggests that the types behavior reported by users follow a pattern that can be exploited in the new reputation system.

Hypotheses 3 and 4 can be verified by an investigation of the leaf categories of the taxonomy. Each of these leaf categories should be sufficiently specific in order to constitute of criterion of behavior evaluation that can be applied by the user (so that the user can evaluate his satisfaction with respect to that criterion). If this hypothesis would not hold, it would not be possible to associate each of the leaf categories with a rating in the proposed new reputation system.

For hypothesis 4 to hold, each leaf category should specify a type of behavior that can be described by a single sentence using modular logic operators such as “must”, “must not”. This hypothesis is not directly required for the design of the new reputation system, but has an interesting theoretical meaning, as it gives empirical support to the notion of normative trust: the definition of trust as an expectation to follow social norms that apply in a certain social context [24].

Hypothesis 5 can be verified if a significant portion of non-positive feedbacks from a large set cannot be classified into a single category of the taxonomy, but can be classified into two or more such categories. If this hypothesis holds, it has an important consequence for the design of the reputation system. Previous work [26] has focused on systems that used multiple criteria that were dependent (increasing the value of one criterion reduced the value of the others). If the criteria are independent, the entire processing of reports by the reputation system should be redesigned.

Hypothesis 6 holds if there exists a preference relation (importance relation) on the criteria for behavior evaluation specified by