Enhanced Recommendations Through Propagation of Trust and Distrust

Patricia Victor, Chris Cornelis, and Martine De Cock
Computational Web Intelligence, Dept. of Appl. Math. and Comp. Sci.
Ghent University, Krijgslaan 281 (S9), 9000 Gent, Belgium
{Patricia.Victor, Chris.Cornelis, Martine.DeCock}@UGent.be

Abstract

The incorporation of a trust network among the users of a recommender system (RS) proves beneficial to the quality and amount of recommendations. Involving also distrust can offer additional clues how to handle specific recommendations as well as protection against recommendation attacks, yet this direction has not been thoroughly explored so far. In this paper, we advocate the use of a trust model for RSs in which trust scores are (trust,distrust)-couples, drawn from a bilattice. We design an experimental setup to get insight into the trust propagation problem in a movie RS and propose two trust score propagation operators, each reflecting a distinct user behaviour pattern or profile.

1. Introduction

Trust models come in many flavours. A *probabilistic* approach deals with trust in a black or white fashion — an agent or source can either be trusted or not — and computes the probability/belief that the agent can be trusted (e.g. [6, 7]). A *gradual* approach is concerned with the computation of trust scores when the outcome of an action can be positive to some extent, e.g. when provided information can be right or wrong to some degree, as opposed to being either right or wrong (e.g. [5, 13]). Note that in real life trust is often interpreted as a gradual phenomenon: humans do not merely reason in terms of 'trusting' and 'not trusting', but rather trusting someone 'very much', 'more or less', etc. Hence, applications in which the agent's abilities are expected to approach the human way of thinking and acting as closely as possible, can highly benefit from this type of models.

RSs match this description very well, and incorporating an appropriate trust model in them can alleviate some major issues such as the cold start and sparsity problem: in [8], it was shown that for new (cold start) users, a few trust statements already yield much higher coverage and reduced error rates. Besides, through propagation of trust values more users (and more products) can be reached, hence alleviating sparsity. Finally, research has shown that people tend to rely upon recommendations from people they trust, more than upon online RSs which generate recommendations based on anonymous people similar to them [12]. Hence, it is obvious that establishing a trust network among RS users can contribute to its success. Some attempts in this direction have been made, most notably Golbeck's work [4].

Because RSs are widely used in the realm of e-commerce, there is a natural motivation for producers of items (manufacturers, authors, etc.) to abuse them so that their items are recommended to users more often [14]. In existing approaches to trust models in RSs, only trusted sources are taken into account. Hence, when a recommending agent is suspected to be malicious, it can be marked as not trusted. However, these approaches do not differentiate between absence of trust caused by presence of distrust (as towards a malicious agent) - versus by lack of knowledge (as towards an unknown agent). In a large recommendation network with many agents who are possibly anonymous and/or unknown to each other, this is a serious drawback.

Therefore, we want to develop a model that is able to make this distinction. Furthermore, we wish to investigate the added value of such a refined model for RSs. To this aim, we first present a provenance-preserving model for trust scores [13] (Sect. 2). To be able to compute with these scores in a trust network embedded in a RS, appropriate propagation operators need to be designed. Section 3 outlines an experiment designed to get insight into this problem and to pinpoint the most common profiles. Its preliminary results yield two operators, each reflecting a distinct profile that can be used to personalize recommendations (Sect. 4). We conclude in Section 5.

2. Trust score space

In [13], we argued that representing trust as a combination of *two values* helps to preserve valuable trust provenance information indicating why problems regarding trust and knowledge arose, viz. (1) absence of trust (value=0) caused by a presence of distrust versus by a lack of knowledge, and/or (2) having too much (inconsistent) or too little information (ignorance). In particular, we proposed a model that treats *partial trust* and *partial distrust* as two different but related concepts. Although it is acknowledged that distrust can play an important role in many applications, besides our own work [1, 13], we are only aware of two other models [5, 6] that take into account both trust and distrust[1]. In [13] we proposed an extension of [1] in which trust values are derived from a bilattice [3] resulting in a new gradual model for (trust, distrust)-couples, called *trust scores*. More details on the rationale behind the model can be found in [13].

Definition 1 (Trust Score Space) *The trust score space*

$$\mathcal{BL}^{\square} = ([0,1]^2, \leq_t, \leq_k)$$

consists of the set $[0,1]^2$ *of trust scores and two orderings defined by*

$$(x_1, x_2) \leq_t (y_1, y_2) \text{ iff } x_1 \leq y_1 \text{ and } x_2 \geq y_2$$

$$(x_1, x_2) \leq_k (y_1, y_2) \text{ iff } x_1 \leq y_1 \text{ and } x_2 \leq y_2$$

for all (x_1, x_2) *and* (y_1, y_2) *in* $[0,1]^2$*. In the trust score* (x_1, x_2)*,* x_1 *is called the trust degree, while* x_2 *is the distrust degree.*

Figure 1 shows $\mathcal{BL}^{\square}$, along with some examples of trust scores. The lattice $([0,1]^2, \leq_t)$ orders the trust scores going from complete distrust $(0,1)$ to complete trust $(1,0)$. The lattice $([0,1]^2, \leq_k)$ evaluates the amount of available trust evidence, ranging from a "shortage of evidence", $x_1 + x_2 < 1$ (incomplete information), to an "excess of evidence", viz. $x_1 + x_2 > 1$ (inconsistent or contradictory information). In the extreme cases, there is no information available: $(0,0)$; or there is evidence that says that the agent is to be trusted fully as well as evidence that states that the agent is completely unreliable: $(1,1)$.

Taking into account a distrust degree next to the traditionally considered trust degree makes it possible to distinguish between absence of trust caused by a lack of knowledge, i.e. $(0,0)$, versus absence of trust in the presence of distrust, i.e. $(0,1)$.

3. Trust score propagation experiment

Trust score propagation is the computation of a meaningful trust score for agent a in agent c, given the trust scores for a in agent b, as well as for b in c. As an agent's trust behaviour depends in most cases on the situation or goal, we focus specifically on propagation in a movie RS and designed an experimental setup according to this scenario. Furthermore, we want to investigate if a provenance-preserving trust model can contribute to the quality and amount of the recommendations. In other words, can information coming from distrusted or unknown agents be used in the recommendation process?

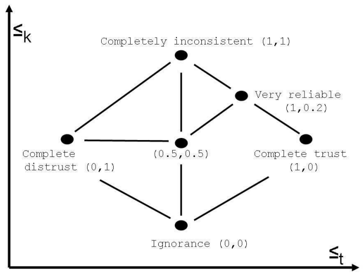


Figure 1. Trust score space $\mathcal{BL}^{\square}$

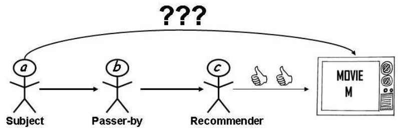


Figure 2. Recommending a movie

Experimental setup Each subject (a) was asked to imagine himself in a movie theatre, having no clue on which movie to see. A passer-by (b) tells him he knows someone (c, the recommender) who has seen movie m and liked it a lot (fig. 2). We distinguish 3 basic types of passers-by: someone you *completely trust* when it comes to movies $(1,0)$, a person you *fully distrust* $(0,1)$, and *a stranger*, a person you have never seen before $(0,0)$. The recommender belongs to one of these types as well. The subjects were then asked how to handle c's advice. An example of such a question is given in fig. 3. We provided six possible answers, with each of the choices/actions accompanied by their intended meaning, in order to exclude as much misunderstandings as possible.

Our goal is to predict the score of a in c. As it may be hard for the subjects to express this score explicitly, we instead ask whether a would follow c's advice. Following Gambetta's well known trust definition [2], a's action/ answer will give us an indication of how much a trusts c.

Preliminary results So far, 24 people have taken part in the experiment[2]. Early results indicate that half of the users present a clear attitude. Two profiles came to the fore, each of them followed by the same amount of people. Subjects with *profile 1* follow the opinion of a trusted agent b: if b distrusts c, the subject would not see the movie, despite

[1] For a discussion of their differences, we refer to [13].

[2] This is an ongoing experiment. Please feel free to participate at `www.cwi.ugent.be/patricia.html`.

the fact that c gave a positive recommendation (hence distrusting c, as b's distrust in c suggested). Besides, these subjects ignore everything coming from a stranger b: irrespective of b's trust in c, the subjects decided not to pay any attention to the recommendation. Finally, they reverse the opinion of a distrusted b: if b distrusts/trusts c, the subject would still/not see m when c gives a positive recommendation (hence trusting/distrusting c). This behaviour can informally be described as 'the enemy of your enemy is your friend', and 'the friend of your enemy is your enemy too', friend (enemy) denoting a person that is (dis)trusted.

Subjects with *profile 2* also follow a trusted b and ignore everything coming from an unknown party. When encountering a distrusted b however, they reverse b's opinion when b trusts c (as in profile 1), but ignore the advice of the recommender when b distrusts c. In other words, the subect does not consider the enemy of his enemy as his friend.

4. Propagation operators

All results indicate that, if only trust is involved, a trusts c if a trusts b *and* b trusts c. In other words, the propagation comes down to the conjunction of the given trust values. To model conjunction in our approach — where trust degrees range from 0 to 1 — we use a triangular norm [11] (t-norm for short) T: an increasing, commutative and associative $[0,1]^2 \to [0,1]$ mapping satisfying $T(x,1) = x$ for all x in $[0,1]$. One can easily verify that $T(0,0) = T(0,1) = T(1,0) = 0$ and $T(1,1) = 1$, hence T is a conservative extension of boolean conjunction. Examples are given in Table 1. The choice of $T = T_{\mathrm{P}}$ corresponds to a common approach in trust propagation (see e.g. [4]).

This procedure is quite straightforward, but when taking into account distrust the picture gets more complicated, as the results of the experiment indicate. They show us that there are multiple possible propagation scenarios for trust scores, differing from person to person. Bearing this in mind, we introduce three propagation schemes[3], or profiles, corresponding to the results of our experiment.

An agent a with profile 1 trusts c when a trusts b *and* b trusts c, *or* when a distrusts b *and* b distrusts c. Furthermore, such an agent a distrusts c when a trusts b *and* b distrusts c, *or* when a distrusts b *and* b trusts c. Analogously to the t-norm, we use a t-conorm S to model disjunction: an increasing, commutative and associative $[0,1]^2 \to [0,1]$ mapping satisfying $S(x,0) = x$ for all x in $[0,1]$, hence S is a generalisation of the classical disjunction (examples in Table 1). This profile is reflected by $\mathtt{Prop}_1$:

Definition 2 (profile 1) *For* (t_i, d_i) *in* $\mathcal{BL}^{\square}$, $i = 1, 2, 3$*:*
$\mathtt{Prop}_1((t_1,d_1),(t_2,d_2)) = (t_3,d_3)$*, with*
$t_3 = S(T(t_1,t_2),T(d_1,d_2))$*,* $d_3 = S(T(t_1,d_2),T(d_1,t_2))$*.*

$T_{\mathrm{M}}(x,y) = \min(x,y)$	$S_{\mathrm{M}}(x,y) = \max(x,y)$
$T_{\mathrm{P}}(x,y) = x \cdot y$	$S_{\mathrm{P}}(x,y) = x + y - x \cdot y$
$T_{\mathrm{W}}(x,y) = \max(x+y-1,0)$	$S_{\mathrm{W}}(x,y) = \min(x+y,1)$

Table 1. Examples of t-norms and t-conorms

$\mathtt{Prop}_1$	$(0,0)$	$(0,1)$	$(1,0)$
$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$
$(0,1)$	$(0,0)$	$(1,0)$	$(0,1)$
$(1,0)$	$(0,0)$	$(0,1)$	$(1,0)$

$\mathtt{Prop}_2$	$(0,0)$	$(0,1)$	$(1,0)$
$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$
$(0,1)$	$(0,0)$	$(0,0)$	$(0,1)$
$(1,0)$	$(0,0)$	$(0,1)$	$(1,0)$

Figure 4. Examples of profile 1 and 2

(t_1,d_1) should be interpreted as the trust score of a in b. Similarly, (t_2,d_2) and (t_3,d_3) are the trust scores from b in c and from a in c respectively. The formulas for t_3 and d_3 in def. 2 involve t_1 as well as d_1, reflecting that a trusted as well as a distrusted passer-by b will be taken into account when deriving the propagated trust score.

Its behaviour for some particular trust scores is shown in fig. 4. The rows/columns correspond resp. to (t_1,d_1)/(t_2,d_2). Such an agent indeed considers an enemy of an enemy to be a friend: $\mathtt{Prop}_1((0,1),(0,1))$=$(1,0)$, and a friend of an enemy to be an enemy: $\mathtt{Prop}_1((0,1),(1,0))$= $(0,1)$. This profile shows us that useful information can be derived through distrusted agents.

An agent with profile 2 is similar to $\mathtt{Prop}_1$, but does not take over information coming from a distrusted agent c when b is distrusted. In other words, c is only trusted by a when a trusts b *and* b trusts c:

Definition 3 (profile 2) *For* (t_i, d_i) *in* $\mathcal{BL}^{\square}$, $i = 1, 2, 3$*:*
$\mathtt{Prop}_2((t_1,d_1),(t_2,d_2)) = (t_3,d_3)$*, with*
$t_3 = T(t_1,t_2)$*,* $d_3 = S(T(t_1,d_2),T(d_1,t_2))$*.*

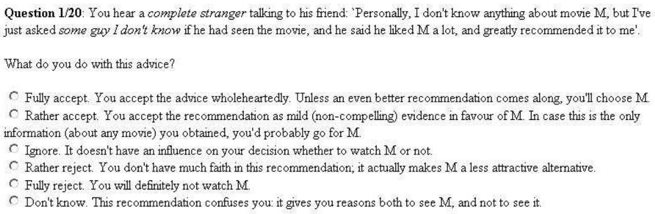
The following example illustrates the effects of partial trust and partial distrust. We choose T=T_P and S=S_P.

Example 1 Although a highly trusts b, there is also evidence to slightly distrust b, e.g. (t_1,d_1)=$(0.8,0.3)$. Agent b highly distrusts c: (t_2,d_2)=$(0.1,0.9)$. We obtain $\mathtt{Prop}_2((t_1,d_1),(t_2,d_2))$=$(0.08,0.7284)$: a takes over most of the information that b provides. However, the final trust score is mitigated because a also slightly distrusts b.

5. Conclusions and future work

Incorporating appropriate trust models proves beneficial to RSs. Although it is acknowledged that distrust can play an important role too, this direction has not been thoroughly explored yet in RSs. In particular, existing approaches do not differentiate between absence of trust caused by presence of distrust (as towards a malicious agent) - versus by

[3] Note that these are not identical to the ones proposed in [13]: some of the operators proposed there were not encountered in any subject, while the results have suggested profile 2, which was not covered by [13].



Question 1/20: You hear a *complete stranger* talking to his friend: `Personally, I don't know anything about movie M, but I've just asked *some guy I don't know* if he had seen the movie, and he said he liked M a lot, and greatly recommended it to me'.

What do you do with this advice?

○ Fully accept. You accept the advice wholeheartedly. Unless an even better recommendation comes along, you'll choose M.
○ Rather accept. You accept the recommendation as mild (non-compelling) evidence in favour of M. In case this is the only information (about any movie) you obtained, you'd probably go for M.
○ Ignore. It doesn't have an influence on your decision whether to watch M or not.
○ Rather reject. You don't have much faith in this recommendation; it actually makes M a less attractive alternative.
○ Fully reject. You will definitely not watch M.
○ Don't know. This recommendation confuses you: it gives you reasons both to see M, and not to see it.

Figure 3. Example of the questionnaire

lack of knowledge (as towards an unknown agent). Experimental results described in this paper indicate that users do tend to make a distinction. More specifically, we found that while unknown parties are largely ignored in trust score propagation, potentially useful recommendation information can be derived through distrusted third parties. Unfortunately, unlike when only taking into account trust, in the presence of distrust, different propagation scenarios are possible. In particular, we encountered both a scenario in which the friend of an enemy is considered an enemy, and an even stronger scenario in which additionally the enemy of an enemy is considered to be a friend (with friend (enemy) denoting a person that is (dis) trusted). We proposed operators reflecting each of these user behaviour patterns.

Although the results of our experiment are preliminary, we argue that by using a provenance preserving trust model, and by explicitly taking into account distrust into the recommendation process, more valuable information becomes available. This can be used to refine other techniques such as collaborative filtering [9] or content based filtering [10], as well as to make these techniques less vulnerable to recommendation attacks. Since the idea of taking distrust into account in RSs is new, much ground remains to be covered, both regarding the problem of trust score propagation as well as subsequent problems. E.g., as the final goal of a RS is to deliver personalized recommendations, how should we combine the existing recommendations with the available trust scores? More specifically, which approach delivers the best results: propagation of trust scores and a one-time combination with the original recommendation at the end, or propagation of recommendations, where each trust score is combined with the recommendation an agent (b) receives from a former agent (c)? Other subsequent problems that need to be addressed include aggregation (combining several propagation chains) and trust score updating.

Acknowledgments

Patricia Victor would like to thank the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) for funding her research. Chris Cornelis would like to thank the Research Foundation-Flanders for funding his research.

References

[1] M. De Cock and P. Pinheiro da Silva. A many-valued representation and propagation of trust and distrust. *LNCS*, 3849:108–113, 2006.

[2] D. Gambetta. Can we trust trust? In *Trust: Making and Breaking Cooperative Relations*, pages 213–237.

[3] M. Ginsberg. Multi-valued logics: A uniform approach to reasoning in artificial intelligence. *Comput Intell*, 4:265–316, 1988.

[4] J. Golbeck. *Computing and applying trust in web-based social networks*. PhD thesis, 2005.

[5] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *Proceedings of WWW2004*, pages 403–412, 2004.

[6] A. Jøsang and S. Knapskog. A metric for trusted systems. In *Proceedings of NIST-NCSC 1998*, pages 16–29, 1998.

[7] S. Kamvar, M. Schlosser, and H. Garcia-Molina. The eigentrust algorithm for reputation management in P2P networks. In *Proceedings of WWW2003*, pages 640–651, 2003.

[8] P. Massa and P. Avesani. Trust-aware collaborative filtering for recommender systems. *LNCS*, 3290:492–508, 2004.

[9] P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm, and J. Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of CSCW 1994*, pages 175–186, 1994.

[10] J. Schafer, J. Konstan, and J. Riedl. E-commerce recommendation applications. *Data Min Knowl Disc*, 5:115–153, 2001.

[11] B. Schweizer and A. Sklar. Associative functions and statistical triangle inequalities. *Publ Math-Debrecen*, 8:169–186, 1961.

[12] R. Sinha and K. Swearingen. Comparing recommendations made by online systems and friends. In *Proceedings of the DELOS-NSF Workshop on Personalisation and Recommender Systems in Digital Libraries*, 2001.

[13] P. Victor, M. De Cock, C. Cornelis, and P. Pinheiro da Silva. Towards a provenance-preserving trust model in agent networks. In *Proceedings of Models of Trust for the Web, WWW2006 Workshop*, 2006. Available at `http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-190/`.

[14] S. Zhang, Y. Ouyang, J. Ford, and F. Makedon. Analysis of a Low-Dimensional Linear Model Under Recommendation Attacks. In *Proceedings of ACM SIGIR 2006*, pages 517–524, 2006.