An Empirical Investigation of Similarity-Driven Trust Dynamics in a Social Network

Yugo Hayashi[1], Victor Kryssanov[2], and Hitoshi Ogawa[2]

[1] Faculty of Library, Information and Media Science, Tsukuba University
1-2, Kasuga, Tsukuba, Ibaraki 305-8550, Japan
hayashiyugo@gmail.com
[2] Faculty of Information Science and Engineering, Ritsumeikan University
1-1-1 Nojihigashi, Kusatsu 525-8577, Japan
{kvvictor,ogawa}@is.ritsumei.ac.jp

Abstract. Presently, people often create and keep lists of other people with similar preferences for hobbies, such as books, movies, music, and food in online social network service systems. Recent studies in recommender systems have shown that the user's data can be used to recommend items based on other users' preferences (e.g. as implemented in amazon.com). To make such systems more effective, there is a need to understand the mechanism of human trust formation. The goal of this study is to develop cognitive models describing the trust formation in social networks. This paper presents results of a controlled experiment conducted to collect human behavior data through a series of trust evaluation tasks.

Keywords: Social Cognition, Social networks, Trust Dynamics, Recommender Systems, Human-Computer Interaction, Conversational Agents.

1 Introduction

Studies in recommender systems have suggested that a user's data can be used to recommend this user items, based on other users' preferences. Although, a relevant question comes up: Can one predict how people may rely on and/or trust opinions of others in their own decision-making? To develop efficient and effective recommender systems, there is a need to understand the mechanism of human trust formation. The paper proposes a new experimental design where computer agents are used to capture the dynamics of trust formation through interactions in a social network.

1.1 Similarity and Trust in a Social Network

Studies in social psychology showed that trust is often related with similarity [1]. Studies in human and computer interaction, on the other hand, showed that users tend to prefer recommendations from friends rather than from computer systems [2]. An experiment conducted by [3] showed that the profile similarity may be related to the

M. Kurosu (Ed.): Human-Computer Interaction, Part III, HCII 2013, LNCS 8006, pp. 20–28, 2013.

ways the users determine whether to trust other users when solving an on-line selection task. The study provided implications for how to predict trust, and how the corresponding model would be incorporated into related algorithms, such as collaborative filtering algorithms used in recommender systems. The study focused on the relationship between similarity and trust but the proposed model does not address the dynamics of user similarity in an online-social network.

Given this background, two relevant questions arise: 1) how could one develop a behavioral model of the similarity and trust development for the application in recommender systems, using information extended from social networks, and 2) can one simply use similarity as an index of trust? The presented study investigates these questions in an open-ended environment, where individuals interact through time.

1.2 Recommendations by Other Users in a Social Network

In social network studies, there were attempts to develop recommender systems and algorithms based on the users' similarity and trust (e.g. [4], [5], and [6]). A popular method to provide the user with adequate recommendations is the so-called collaborative filtering. Collaborative filtering allows for recommending items to users, based on their similarity and/or on the similarity of the items the users prefer to other items. A more advanced method for recommending items to the users is the Bayesian filtering algorithm. Using the latter algorithm, it is often possible to predict the dynamics of human trust, based on the past results of trust/distrust and similar/dissimilar evaluations. There were, however, few studies that would try to validate the Bayesian framework with real human data. As many studies in psychophysics showed, psychological states related to trust may differ due to differences in physical world. To address this problem, the presented study focuses on an experimental situation where an individual perceives a single dissimilar physical state while interacting with other users.

1.3 This Studies Objective

This study thus aims to explore how human participants react towards the recovery of trust throughout the interaction process. It investigates differences between a simulation model currently used in many recommender systems and the corresponding human performance observed in an experiment. The commonly used Naïve Bayes model is validated with empirical data as a correct model of the human trust dynamics. Data generated with the Bayes model is used as the criterion for comparison with empirical human data for understanding parameters of the trust development in a social network. The following three issues are examined in the study:

1. How similar/dissimilar preferences affect trust dynamics in an (online) social network?

2. Assuming that after some time, similar preferences may contribute to the development of a high trust and a relationship, then when one perceives a dissimilar state, what is the immediate dynamics of trust?
3. How would the empirical data differ with the corresponding normative behavioral data obtained through simulation with the classical Naïve Bayes model?

2 Method

A web-based system has been developed for conducting a controlled experiment to understand cognitive aspects of trust in a social network. In the proposed experimental paradigm, a multi-conversational agent was used as an imaginary actor playing the role of a member of the social network. This experimental method helped us understand specific stimulatory responses, as observed in trust dynamics, when an unexpected incidence occurs in the social network.

2.1 Procedures

In the experiment, participants (who are all members of one social network) evaluated opinions (preferences) of other participants for a topic provided. More specifically, they watched a video-clip of a short cartoon, using smart phones, and shared their impressions about the cartoon with other participants through a web-page. The procedure was as follows: 1) Participants first watch the video, using smart phones, 2) They evaluate how interesting they felt about it, 3) They see the other users' evaluation results and, finally, 4) They evaluate how strong they 'trust' the other users (see Fig. 1).

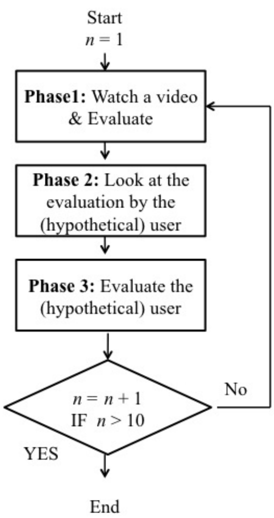

Fig. 1. The experimental procedure

The procedure was continued for about 10 trials (the exact number of trials depended on the experiment). During the task, the participants watched 10 different video clips in a predefined order, evaluated their impressions, and shared the evaluation results with each user in every trial. Results from users were manipulated by computer agents (i.e. by the hypothesized users), and were controlled to change over time. This methodology is adopted from a previous study [7]. One of the hypothesized users' evaluations was adjusted to be always almost the same as the participant's, excepting for one trial where the participant experienced a "shock" event due to completely different evaluations learned from the other users. Also, participants were asked to do a secondary evaluation by evaluating the other users, based on how strong they "trusted" the other users' opinions. Details of the experiment are discussed in the following sections.

Phase 1. First, participants accessed the Web-based application using their mobile devices. In beginning of the task, all participants received the following instruction: "You are randomly connected to four persons in this experiment, who are also now watching the cartoon and giving their evaluations." Actually, there were, however, no other real users but computer conversational agents responding appropriately to each participant. This instruction made participants to think that they interact with members of the social network in real time. The participant starts the task by rating on a collection of short films by accessing You Tube (`http://www.youtube.com`). Episodes of a short cartoon "Tom and Gerry" were used as the stimuli. After the participant finished watching one clip he or she evaluated it, based on three questions (see Fig. 2). Each evaluation was on a ten-point lickert-scale, and the three questions were: "how good was the content", "how good was the character", and "how good was the scenario" (as translated from Japanese).

Fig. 2. Screenshots from the ratings phase

Phase 2. After the participant finished rating, he or she can see how other users rated the episode. All other users' ratings were generated by computer agents. These latter ratings will be called hypothesized users' "rating profiles". Fig. 3 shows an example of the rating profiles.

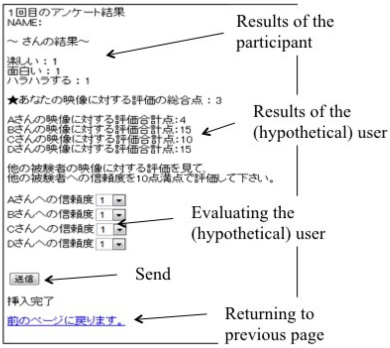

Fig. 3. Screenshots from the rating phase

Two types of the results were shown: (1) the sum of ratings by the participant and (2) the sum of the rating profiles by the hypothesized users. Let us denote the participant's sum of the ratings of the episode in trial i as Si, the hypothesized user's sum of rating profiles as U_a, U_b, U_c, U_d, and the absolute difference between any two ratings as $ß$. For each episode i, $ß$ was randomly selected from the following three categories by the computer agents:

- Small variation : $0 \le ß \le 2$,
- Medium variation: $3 \le ß \le 5$,
- Large variation: $6 \le ß \le 9$.

To investigate how similar/dissimilar preferences affect the trust dynamics, one of the four hypothesized members labeled "User A" was always adjusted to generate small variations. Two other hypothesized members (labeled "User B" and "User C") always generated large variations, and one such member generated medium variations ("User D"). If people trust more to the users with similar ratings, it would be natural to expect that a subject would develop higher trust towards "User A" than to the other users. It also appears natural to expect that this tendency would increase over time.

Table 1. Rating profiles of users and trials

Hypothetical user	Trial i	$ß$
U_a	1, 2, 4, 6, 7, 8, 9, and 10	Small variations
	3, 5	Large variations
U_b	All trials	Large variations
U_c	All trials	Medium variations
U_d	All trials	Large variations

User A was manipulated almost in the same manner as the participant behaved, excepting for trial 3 and 5. In that trial, User A response, U_a, was adjusted to change from Small variations to Large variations. Through this adjustment, the participant experienced a shock, when the result of the evaluation changed dramatically compared to other trials. U_b and U_d were adjusted to change always in Large variations, and U_c was adjusted to change always in Medium variations. The data of the secondary evaluation of trust was collected and used as the main data for the analysis.

Phase 3. In this phase, the participants were asked to rate how much they trusted the hypothetical users, based on what they observed in the rating profiles. Participants evaluated four hypothetical users on a ten-point lickert scale. After finishing this activity, they proceeded to the next trial, and watched a new video clip. This cycle continued until the participants finished 10 trials.

2.2 Participants and Condition

Two experiments were conducted to investigate the cognitive aspects of trust in a preference evaluation task. In both experiments, all the participants were students enrolled in a psychology class. The participants were asked to participate in the experiments for course credits. The experiments took place in the classroom and the participants were told to use their smartphones in the experiments. In the experiment, twenty-seven participants (male = 7, female= 20) made evaluations in ten trials.

3 Results

3.1 Overall Result

To investigate the trust dynamics for a longer period of time, Experiment 2 was conducted. Results obtained in this experiment were then analyzed, using a 4 (users) x 10 (trials) within-subjects factorial design. Fig. 4 shows the results of the trust evaluations for the hypothetical users. The vertical axis gives the average of the trust evaluations, and the horizontal axis represents the trials.

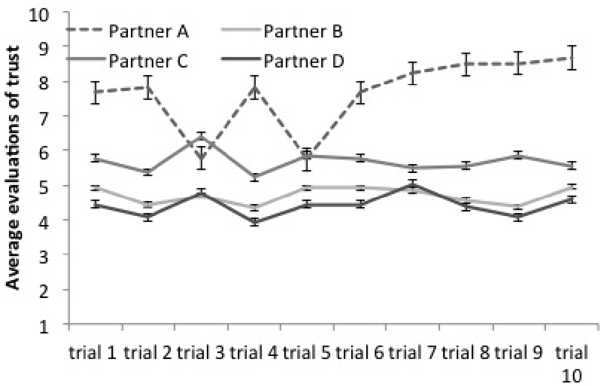

Fig. 4. Results of evaluations

The ANOVA analysis revealed that there was an interaction between the two factors ($F(26,702)=4.768$, $p < .01$). Simple main effects in the users detected several differences among trials. Differences were found in all trials ($F(3,936)=19.753$, $p < .01$; $F(3,936)=27.758$, $p < .01$; $F(3,936)=6.624$, $p < .01$; $F(3,936)=29.880$, $p < .01$; $F(3,936)=4.430$, $p < .01$; $F(3,936)=25.505$, $p < .01$; $F(3,936)=24.200$, $p < .01$; $F(3,936)=35.359$, $p < .01$; $F(3,936)=40.421$, $p < .01$; $F(3,936)=33.805$, $p < .01$). Multiple comparisons, using Ryan's method showed that User A was evaluated higher than users B, C and D in trial 1 ($p < .01$; $p < .01$; $p < .01$). These results were consistent for trials 1, 2, 4, 6, 7, 8, 9, and 10. The result of the two experiments has, therefore, demonstrated that the ratings of trust become higher when the participants observe a similarity to their own opinions in the shared opinions about the cartoon. This finding generally confirms that similar preferences enhance trust. The results obtained also showed that the ratings of trust dropped as the participants experienced the shock of non-similar evaluations. This confirms that similarity of preferences and trust are strongly correlated, and this correlation can be enhanced over time.

3.2 Simulation

A further analysis was conducted, using the Bayesian model for a comparison with the data. Evaluations towards User A were used for investigating how trust recovers after experiencing the shock. In the analysis, evaluations of User A were binary coded to either, "trust" (1) or "no trust" (0). After the second trial, when participants rated lower than in the previous trial, the evaluating were coded as "not trust". When participants rated higher or the same, compared to the previous trial, the evaluations were coded as "trust". The ratio of selecting "trust"/"not trust" was calculated for each trial. Using this coding scheme, all the data was then compared with the Bayesian model. For the analysis, prior probability of perceiving a similar /dissimilar evaluation in trial i was determined as H_i, and the probability of generating trust was $P(H_i|D)$. The following equation specifies the Naïve Bayesian model used:

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{P(D|H_1)P(H_1)+P(D|H_2)P(H_2)} \tag{1}$$

$$where\ \mathrm{i} = 1,2. \tag{2}$$

Fig. 5 shows the empirical data in comparison with simulation results. The vertical axis gives the probability of trust, and the horizontal axis represents the trials. The results demonstrate that the Naïve Bayes model fails to accurately predict the speed of recovering from distrust to trust but still qualitatively is in a good agreement with the observed empirical dynamics.

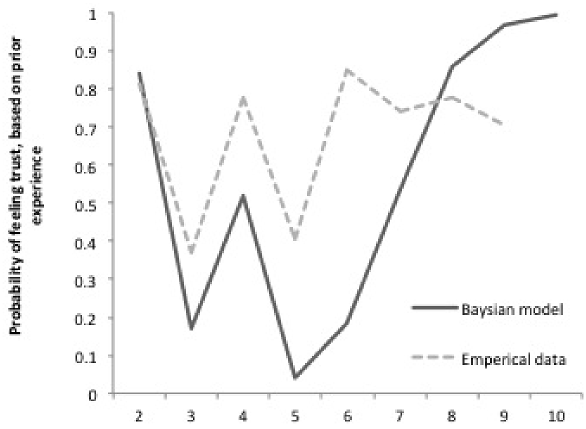

Fig. 5. Results of the simulation

4 Discussion and Conclusions

The trust dynamics observed in the experiments revealed that when one perceives a dissimilar state, the subjective perception of trust decreases temporary, but then almost immediately recovers upon a positive experience. This fact is interesting in the view of understanding the users of various social network service systems, who tend to ignore social conflicts and keep developing their trust in respect to a particular member, while ignoring negative experience of interactions with that member. This phenomenon would be attributed to so-called conformation bias. Several studies have recently shown that people are likely to become biased to (mis)trust others' opinions in an online environment (e.g. [8]). It follows from the results obtained in our study that such bias on trust may rapidly develop in social network service systems. This finding would have implications for design of recommender systems based on information extracted from social networks. In future work, we plan to investigate in detail the relationships connecting trust and similarity to find ways for controlling the conformation bias.

References

1. Byrne, D.: The attraction paradigm. Academic Press, New York (1971)
2. Sinha, R., Swearingen, K.: Comparing recommendations made by online systems and friends. In: Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries. Dublin, Ireland (2001)
3. Golbeck, J.: Trust and Nuanced Profile Similarity in Online Social Networks. ACM Transactions on the Web 3(4-12), 1–32 (2009)
4. Walter, E.F., Battiston, S., Schweitzer, F.: A model of a trust-based recommendation system on a social network. Journal of Autonomous Agents and Multi-Agent Systems 16(1), 57–74 (2007)

5. Ziegler, C.N., Lausen, G.: Spreading activation models for trust propagation. In: Proceedings of the IEEE International Conference on e-Technology, e-Commerce, and e-Service, pp. 83–97. IEEE Computer Society Press, Los Alamitos (2004)
6. Ziegler, C.N., Golbeck, J.: Investigating Correlations of Trust and Interest Similarity. Decision Support Systems 42(2), 460–475 (2006)
7. Hayashi, Y.: The effect of "Maverick": A study of Group Dynamics on Breakthrough in Collaborative Problem solving. In: Proceedings of the 34th Annual Conference of the Cognitive Science Society, pp. 444–449. Lawrence Erlbaum Associates, Hillsdale (2012)
8. Joinson, A.N.: Understanding the psychology of internet behavior: virtual worlds, real lives. Palgrave, Basingstoke (2003)