

Trust Modeling for Opinion Evaluation by Coping with Subjectivity and Dishonesty

Hui Fang

School of Computer Engineering
Nanyang Technological University, Singapore
hfang1@e.ntu.edu.sg

1 Research Program

Our research is within the subfield of modeling trust and reputation in multi-agent systems for online communities. Specifically, in an online community involving users and entities, users provide opinions (ratings) to entities. For each user, we are interested in addressing two problems:

- how to accurately model the reputation of entities by aggregating opinions from all the users (advisors);
- how to cope with the dishonesty of an advisor in providing opinions as well as her subjectivity difference with the user.

To explain, in open online communities, it is not possible for a user to have experience with every entity. Besides, while being advertised perfectly, an entity might turn out to be malicious in fact. Therefore, there is a significant risk for users when selecting an entity among many alternatives. Users then usually rely on the experience or knowledge of other users (advisors), to choose with which entities to interact. However, the quality of opinions from advisors may vary for two reasons. Firstly, some advisors might be dishonest due to their intrinsic natures. They tend to *intentionally* provide overly positive or negative opinions for some entities which are completely contradict with their real experience. Secondly, advisors might be honest, but are *subjectively different* from users. They provide true opinions based on their experience, which might be *unintentionally* misleading for users due to their salient subjectivity difference with users. Thus, it is important for users to evaluate the quality of advisors' opinions in order to determine how much to rely on.

2 Progress to Date

Up to date, we have proposed: 1) a subjectivity alignment approach (called SARC) by considering the subjectivity difference among users; 2) a novel trust model stemmed from the diffusion theory in social science (called DiffTrust) to model advisors' trustworthiness; and 3) a probabilistic graphical trust model (called PGTM) to evaluate the opinions of advisors by separately considering advisors' dishonesty and their subjectivity difference with users.

2.1 Subjectivity Alignment Approach

Subjectivity difference may come from two sources by analyzing the scenario of a user providing a rating (opinion) to-

wards an entity from both psychological and behavioral perspectives: 1) *intra-attribute subjectivity*: the subjectivity in evaluating the same attribute. For example, an entity may be *expensive* for user *a*, while *not so expensive* for user *b*; and 2) *extra-attribute subjectivity*: the subjectivity in evaluating different attributes. For example, a user with better economic conditions may consider an entity's *quality* more heavily, while another user with worse economic conditions may concern more about the *price* of the entity. These two aspects together contribute to the subjectivity difference among users. Due to the subjectivity difference, it may not be effective if a user directly takes other users' ratings towards entities. To address this issue, we have proposed our subjectivity alignment approach (SARC) [Fang *et al.*, 2012] for reputation computation.

In our approach, each user is equipped with an intelligent agent. At the beginning of her interactions with the reputation system, a user *a* is required to provide her agent with both a single rating and a detailed review containing values of the objective attributes of her interactions with entities, such as *price* and *delivery time*, for each of a few interactions. Based on these rating-review pairs, the agent applies a proposed Bayesian learning approach to model the correlations between *a*'s each rating level and the value of each objective attribute involved in the interactions. The learned correlation function, which represents *a*'s *intra-attribute subjectivity*, will then be shared with the agents of other users. The agent of *a* also applies a regression analysis model to learn the weight of each attribute for *a*, representing her *extra-attribute subjectivity*. This information will not be shared with other users. After the learning phase, *a* only needs to provide ratings for her interactions with entities, not detailed reviews. When another user *b* shares a new rating of her interaction with an entity, the agent of *a* will first retrieve a rating level for each attribute of the experience based on the shared rating and the *intra-attribute subjectivity* of *b* shared by the agent of *b*. The rating levels of the attributes will then be aggregated according to *a*'s *extra-attribute subjectivity* learned by the agent of *a*. In this way, the rating shared by *b* is aligned to that can be used by *a*. Experimental results demonstrate that: 1) SARC performs better than the representative competing approaches, and is capable of coping with environments with dynamic user and entity behavior; 3) the requirement of detailed reviews and objective attributes is not very restrictive.

2.2 Trust Model Stemmed from Diffusion Theory

Trust has been recognized as a diffusive concept. When modeling trust, it is crucial to consider the processes through which trust is cultivated in a system. The diffusion theory [Strang and Tuma, 1993] in social science seeks to explain *how*, *why* and *at what rate* a new *innovation* spreads through a community. It is thus natural to derive a trust model from this well-studied theory by considering an advisor's trustworthiness as an innovation [Fang *et al.*, 2013b]. Specifically, an advisor's trust building among users is considered as a diffusion process. Her trustworthiness perceived by a specific user is influenced by: the advisor's characteristics directly observed by the user, susceptibility of the user, the contagious influence of other users already having a certain level of trust on the advisor, and information about the environment (including both spatial and temporal information). With this model, we can well capture the *dynamics* and *subjectivity* of trust, and its *dependency* on other users and the environment. Besides, it can also help model the trustworthiness of advisors for users who are new to the system.

Computationally, each user in the system is equipped with a software agent. Each agent first computes its user's direct trust on every other user (advisor) based on their shared interactions. A shared interaction here means that the user and the advisor have previously interacted with a same entity. The direct trust information will be used to form a social (trust) network for all users. Then when a user encounters an advisor's opinions, the user's agent will compute the trustworthiness of the advisor by incorporating the user's direct trust on the advisor in the social network, the user's susceptibility to trust (i.e. the initial trust assigned to the advisor by the user), and trust evaluation on the advisor computed by the agents of other users in the social network. The agent of the user considers another user's trust evaluation on the advisor by also computing the social proximity of that user with respect to its own user, which indicates how much the agent's own user can rely on that user's trust evaluation. In the model, the trustworthiness of an advisor in the view of each user is measured under a specific context, and we assume that there is a finite set of contexts in a specific system. We also take into account both temporal and spatial information in trust computation. The experimental results on four real datasets demonstrate that our model can more accurately model the trustworthiness of advisors than state-of-the-art approaches in both loosely-connected and well-connected environments.

2.3 Probabilistic Graphical Trust Model

For the two previous models, SARC only considers subjectivity difference between users and advisors, while DiffTrust mainly deals with advisors' dishonesty, and cannot accurately distinguish these two factors. As indicated, dishonesty is an intrinsic property of advisors, while subjectivity difference exists between users and advisors. For a same dishonest advisor, different users may still have different perceptions on her trustworthiness because the subjectivity difference varies between the advisor and each different user. It is thus worthwhile to consider them separately.

Hence, we propose a novel probabilistic graphical model [Fang *et al.*, 2013a] that explicitly distinguishes the

factors of dishonesty and subjectivity difference. Specifically, in an online community, some users may explicitly identify their trust towards some advisors. Given information about ratings and trust relationships (if any), we model the factors of advisors' intrinsic nature (i.e. *benevolence*, *integrity* and *competence* [McKnight and Chervany, 2001]), users' *propensity* to trust advisors, and *subjectivity difference* between users and advisors, as latent variables in the model that may influence users' trust towards advisors. Through extensive experiments on three real datasets, it is confirmed that our model largely outperforms competing approaches for modeling advisor trustworthiness.

3 Future Research

For the future, we plan to extend our current trust models. More specifically, our current SARC approach works well if there is only one context in the system (e.g., where there is only one type of entities). However, users may have different subjectivity for evaluating different kinds of entities (or the attributes of the different kinds of entities). In this case, we may extend our current approach by taking into account the correlation between the attributes of these different kinds of entities. Our current DiffTrust model focuses on modeling the trustworthiness of advisors. We will derive a trust model also from the diffusion theory to model the trustworthiness (reputation) of entities to which the advisors provide opinions. Together, we will offer a unified trust model for users to make informed decisions about which entities to interact with. Moreover, we will extend our current PGTM model to address other scenarios (e.g., multi-nominal degrees of trust other than binary case), and further verify its effectiveness in other applications (e.g., recommender systems).

We consider to combine the PGMT (or DiffTrust) model with our SARC approach by proposing a two layered approach, where in the first layer, opinions from dishonest advisors are filtered out by using the PGMT (or DiffTrust) model, and in the second layer, opinions from advisors with subjectivity difference are aligned by adopting the SRAC approach.

References

- [Fang *et al.*, 2012] Hui Fang, Jie Zhang, Murat Sensoy, and Nadia Magnenat-Thalmann. SARC: subjectivity alignment for reputation computation. In *Proceedings of the AAMAS*, pages 1365–1366, 2012.
- [Fang *et al.*, 2013a] Hui Fang, Yang Bao, and Jie Zhang. Misleading opinions provided by advisors: Dishonesty or subjectivity. In *Proceedings of the IJCAI*, 2013.
- [Fang *et al.*, 2013b] Hui Fang, Jie Zhang, and Nadia Magnenat-Thalmann. A trust model stemmed from the diffusion theory for opinion evaluation. In *Proceedings of the AAMAS*, 2013.
- [McKnight and Chervany, 2001] D. Harrison McKnight and Norman L. Chervany. What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology. *International Journal of Electronic Commerce*, 6(2):35–59, December 2001.
- [Strang and Tuma, 1993] David Strang and Nancy Brandon Tuma. Spatial and temporal heterogeneity in diffusion. *American Journal of Sociology*, 99(3):614–639, 1993.