Misleading Opinions Provided by Advisors: Dishonesty or Subjectivity

Hui Fang Yang Bao† Jie Zhang
School of Computer Engineering, Nanyang Technological University, Singapore
† School of Computing, National University of Singapore, Singapore
{hfang@e.ntu.edu.sg, baoyang@comp.nus.edu.sg, zhangj@ntu.edu.sg}

Abstract

It is indispensable for users to evaluate the trustworthiness of other users (referred to as *advisors*), to cope with possible misleading opinions provided by them. Advisors' misleading opinions may be induced by their *dishonesty*, *subjectivity difference* with users, or both. Existing approaches do not well distinguish the two different causes. In this paper, we propose a novel probabilistic graphical trust model to separately consider these two factors, involving three types of latent variables: *benevolence*, *integrity* and *competence* of advisors, *trust propensity* of users, and *subjectivity difference* between users and advisors. Experimental results on real datasets demonstrate that our method advances state-of-the-art approaches to a large extent.

1 Introduction

In large, open online communities, users often encounter entities (e.g. people, products and information) which they have no previous experience with or prior knowledge of. To choose which entities to interact with, they usually rely on the experience or knowledge of other users (*advisors*). However, advisors might provide misleading opinions for two very different reasons. Firstly, some advisors might be dishonest due to their intrinsic natures such as the low level of *benevolence*, *integrity* and *competency* [McKnight and Chervany, 2001]. 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* [Fang *et al.*, 2012] 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.

It is thus important for users to evaluate the quality of advisors' opinions in order to determine how much to be relied on. One generally adopted approach [Teacy *et al.*, 2006] is to model the trustworthiness of advisors. The basic idea is that a more trustworthy advisor to a user will provide higher quality opinions to the user. However, some existing trust models may only consider either advisors' dishonesty [Zhang and Cohen, 2007] or subjectivity difference [Fang *et al.*, 2012], while others cannot accurately distinguish these two different factors [Regan *et al.*, 2006; Noorian *et al.*, 2011]. As indicated, dishonesty is an intrinsic property of advisors, while subjectivity difference exists between users and advisors. Even for a same (dis)honest advisor, different users may have different perceptions on her trustworthiness due to the different level of subjectivity difference between the advisor and each user. It is thus necessary to clearly distinguish these two factors.

In this paper, we propose a novel probabilistic graphical trust model that explicitly distinguishes the factors of dishonesty and subjectivity difference. Specifically, in an online community involving users, advisors and entities, both users and advisors can provide ratings to entities. Some users may explicitly identify their (dis)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*), 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 detailed experiments on three real datasets, it is confirmed that our model largely outperforms competing approaches for modeling advisor trustworthiness.

2 Related Work

Different approaches have been proposed to model advisor trustworthiness for users. Some approaches, such as [Teacy *et al.*, 2006; Zhang and Cohen, 2007; Liu *et al.*, 2011], focus on low quality opinions (unfair ratings) intentionally provided by dishonest advisors. Due to the ignorance of subjectivity difference between advisors and users, they may misuse some important information caused by subjectivity difference, rather than dishonesty. Some other approaches, such as [Fang *et al.*, 2012], only consider subjectivity difference between users and advisors, but ignore the dishonesty characteristic of advisors. They may mistakenly treat dishonest advisors as those having subjectivity difference with users.

To model both dishonesty and subjectivity of advisors, BLADE [Regan *et al.*, 2006] and Prob-Cog [Noorian *et al.*, 2011] are proposed recently. BLADE applies Bayesian learning to model the correlations between entities' properties and ratings of users and advisors. However, it does not explicitly distinguish dishonesty and subjectivity in its modeling process. If the correlations learned for users' ratings are based

on entities' properties that are different from those for advisors, it is likely that advisors having subjectivity difference are treated as dishonest. Prob-Cog is a two-layered behavioral modeling approach. First, it filters dishonest advisors according to the rating similarity between users and advisors. Second, trustworthiness of honest advisors is discounted according to their subjective trends. However, Prob-Cog has the assumption that advisors providing very different ratings with a user are dishonest to the user. In consequence, advisors having large subjectivity difference with the user will be misclassified as dishonest. In contrast, our model explicitly distinguishes (dis)honesty and subjectivity difference by modeling them using different sources of rating information, and captures their relationships with trust through the influence of chains in our probabilistic graphical model.

Our model also adopts the trust topology widely studied in Social Science [McKnight and Chervany, 2001] to model advisor (dis)honesty. A similar model is TAF of Chua and Lim [2010], which considers entities' competency, and users' trust propensity and contingency. TAF targets at a different research problem that models users' trust towards entities in online communities. We model advisor trustworthiness and additionally consider *benevolence* and *integrity* of advisors.

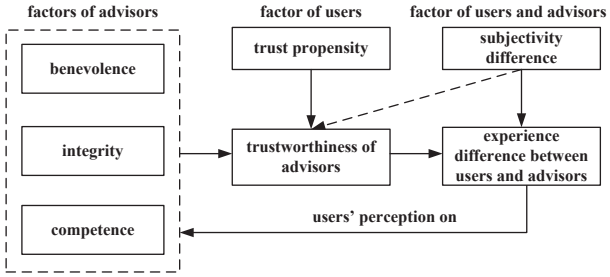


Figure 1: The Conceptual Framework of Trust

3 The Probabilistic Graphical Trust Model

We first identify the factors that influence users' trust towards advisors, to construct a conceptual framework of trust shown in Figure 1. More specifically, in Social Science, *benevolence*, *integrity* and *competence* are regarded as antecedents of trust [McKnight and Chervany, 2001] to explore the relationship between a trustor (user) and a trustee (advisor). They are the intrinsic characteristics of advisors' (dis)honesty that may directly influence users' trust towards advisors (drawn as solid lines). An intrinsic characteristic of users–*trust propensity*, referring to users' initial trust towards unknown advisors before interacting with them, may also have direct effect on users' perception on advisor trustworthiness [Mayer *et al.*, 1995]. In reality, users' different background will conduce to their different initial trust towards others. Different users may have different subjectivity in evaluating same entities. Even an advisor is honest, a user may have different experience with the same entity compared to what the advisor has. In other words, *subjectivity difference* between a user and an advisor can affect the user's satisfaction level of the opinions provided by the honest advisor, and further indirectly influence the user's perception on the advisor's trustworthiness (drawn as the dashed line). Furthermore, experience difference between the user and the advisor, as well as that between the advisor and other advisors (users) can have direct impact on the user's perception towards the advisor's benevolence, integrity and competence.

In the next subsections, we first show the probabilistic graphical trust model designed according to the conceptual framework. We then present its parameters and the generative process of sampling observable variables. We also elaborate the inference process of the model and estimation of parameters. Finally, we predict advisor trustworthiness for users.

3.1 Parameters and Generative Process

In some online communities, e.g. Epinions (`epinions.com`), users may explicitly indicate their trust towards some advisors (trust links), while in others, e.g. eBay (`ebay.com`), no trust links are available. We thus design two graphical models, shown in Figures 2(a) and 2(b), for the two types of communities, respectively. We use graphical model mainly because it can fully interpret our conceptual framework, and seamlessly merge supervised and unsupervised learning (labeled and unlabeled relationship) [Jordan *et al.*, 1999].

We assume a user u and an advisor a in a community, and denote a's trustworthiness perceived by u as $t_{u,a} \in [0, 1]$ where 1 means totally trust and 0 totally distrust[1]. Some other parameters are: 1) u's trust propensity y_u; 2) a's competence c_a, benevolence b_a and integrity i_a; and 3) a's subjectivity difference with u, $s_{u,a}$. All these parameters are modeled as distribution parameters, and the expected values are in $[0, 1]$. Specifically, the expected *trust propensity* y_u of 1 represents complete propensity to trust, while that of 0 no propensity.

Competence (c) indicates whether an advisor has ability to provide reliable ratings (opinions) to users. Hence, it is reasonable to regard it as a latent variable that directly connects with the trustworthiness of the advisor. Its expected value 1 refers to full competence, while 0 means no competence.

Benevolence (b) refers to the degree that an advisor cares about the preferences of users [McKnight and Chervany, 2001]. Thus, we can easily observe its relationship with the rating difference (r) towards the same entity between every user and the advisor. The higher benevolence of the advisor with respect to users will lead to the smaller rating difference between every user and the advisor. Through the chains in the graphical models (see Figure 2, where r is observable), we capture the relationship of benevolence b with advisor trustworthiness t. The benevolence of advisor a consists of two components, one for trust and another for distrust denoted by $b_{a|t=1}$ and $b_{a|t=0}$ respectively.

Integrity (i) emphasizes on the degree that a person follows rules in an organization [McKnight and Chervany, 2001]. Then, it can be inferred that integrity affects rating difference (R) between the advisor and average of all other users for the same entity. Higher integrity implies smaller rating difference. Similar to benevolence, we model the relationship

[1]In model inference, users' trust towards advisors is labeled as either $t = 1$ (trust) or $t = 0$ (distrust) with respect to a predefined threshold, and its corresponding probability value refers to the exact trustworthiness of advisors perceived by each specific user.

of integrity with advisor trustworthiness (see Figure 2, where R is observable), and $i_{a|t=1}$ and $i_{a|t=0}$ represent a's integrity for trust and distrust respectively.

Subjectivity difference $s_{u,a}$ between user u and advisor a may directly influence rating difference (r) between u and a. Through the chains in the models, we seize its influence on trust modeling. Similarly, subjectivity difference also consists of two components, one for trust and another for distrust denoted by $s_{u,a|t=1}$ and $s_{u,a|t=0}$ respectively.

In addition, for communities where trust links are partially observable, we identify a new latent parameter called *expressiveness* denoted as e_u, representing user u's tendency to express her trust $e_{u|t=1}$ or distrust $e_{u|t=0}$ links towards advisors. Note that Figure 2(b) is a special case for Figure 2(a), of which the expected expressiveness e_u always equals to 0.

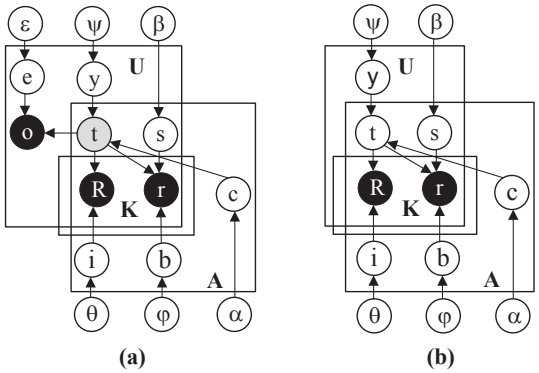


Figure 2: Graphical Model: (a) with (b) without Trust Links

Figure 2 shows dependencies among variables and parameters. The directed arrows represent dependency relationships. Variables in shaded circles (e.g. $r_{u,a,k}$ and $o_{u,a}$) denote observable variables, while the remaining variables are not observable. The variable in grey circle (i.e. $t_{u,a}$) is partially observable in Figure 2(a). The boxes are "plates" representing replicates, where A represents all advisors, U all users and K all entities, respectively. The generative process of our model follows the steps below:

Step 1: For each user $u \in U$, sample distribution parameter of propensity y_u, using Beta distribution with symmetric hyper-parameters, as $y_u \sim Beta(\psi)$.

Step 2: For each advisor $a \in A$, sample distribution parameters: competence c_a, benevolence $b_{a|t}$ and integrity $i_{a|t}$ using Beta distribution with symmetric hyper-parameters, as $c_a \sim Beta(\alpha)$, $b_{a|t} \sim Beta(\varphi)$, and $i_{a|t} \sim Beta(\theta)$.

Step 3: For each user $u \in U$ and each advisor $a \in A$: **[3.1]** u generates trust for a, $t_{u,a}$ (0 or 1), based on u's propensity y_u and a's competence c_a, as $t_{u,a} \sim Bern(y_u) \cdot Bern(c_a)$; **[3.2]** u samples distribution parameters: expressiveness $e_{u|t}$ and subjectivity difference $s_{u,a|t}$ using Beta distribution with symmetric parameters, as $e_{u|t} \sim Beta(\varepsilon)$ and $s_{u,a|t} \sim Beta(\beta)$; **[3.3]** u generates observability of link $o_{u,a}$ based on u's expressiveness e_u and trust of the link $t_{u,a}$, as $o_{u,a} \sim Bern(e_{u|t})$; **[3.4]** for each entity $k \in K$, generate a's rating difference with u, $r_{u,a,k}$, and a's rating difference with all the other users, $R_{U_{-a},a,k}$, as $r_{u,a,k} \sim Bin(m, 1 - b_{a|t}) \cdot Bin(m, s_{u,a|t})$ and $R_{U_{-a},a,k} \sim Bin(m, 1 - i_{a|t})$, where $R_{U_{-a},a,k}$ and $r_{u,a,k}$ are in $[0, m]$.

We use Binomial distribution [Chua and Lim, 2010] to model rating difference (r and R). The basic idea is that, by dividing an entity into m parts, we obtain rating difference of m if the difference for every part is 1, $m-1$ if that for every part is 1 except one part, and 0 if that for every part is 0. We model each part as a Bernoulli event, and then the rating difference can be generated as Binomial distribution. We choose Binomial distribution instead of multinomial distribution mainly because: 1) the assumption of multinomial distribution (each state is independent and identically distributed) does not hold here; and 2) Binomial distribution fits our problem well since r and R are in a finite range.

Also note that there are two kinds of trust links between users and advisors (*trust* and *distrust*). We thus choose Beta distribution with symmetric hyper-parameters for parameters, where each expected value of these parameters from prior distribution is assumed to be 0.5.

3.2 Model Inference and Parameter Estimation

We use Gibbs sampling [Casella and George, 1992] to infer our two models (Figures 2(a) and 2(b)), and then update posterior distribution for each parameter, correspondingly.

Model Inference

Gibbs sampling, as a means of approximate inference, is easy to derive for Bayesian inference (our research case), and comparable in speed to other estimators (e.g. EM algorithm). It is well-adapted to sample posterior distribution of a Bayesian network and approximate a global maximum. Because the conjugacy of Beta and Binomial distributions, we apply collapsed Gibbs sampling [Liu, 1994], which integrates the parameters in Figure 2, for model inference.

Specifically, when trust links are partially observable, as shown in Figure 2(a), we conduct sampling whenever we encounter a trust link of $t_{u,a}$ that is needed to be identified. The Gibbs sampling inference process is:

$$\begin{aligned} &P(t_{u,a} = t | t_{u,-a}, o, r, R, \alpha, \varphi, \theta, \beta, \psi, \varepsilon) \\ &\propto P(t_{u,a} = t | t_{u,-a}, \alpha, \psi) P(R_{U_{-a},a} | t_{u,a} = t, R_{U_{-a},-a}, \theta) \\ &\quad P(o_{u,a} | t_{u,a} = t, o_{u,-a}, \varepsilon) P(r_{u,a} | t_{u,a} = t, r_{u,-a}, \varphi, \beta) \end{aligned}$$

where $t_{u,-a}$ refers to user u's trust links with other advisors except advisor a; $o_{u,-a}$ represents the observability of u's trust links except that with a; $R_{U_{-a},-a}$ is the rating difference between each of other advisors (except advisor a) and the average rating; and $r_{u,-a}$ is the rating difference between user u and each of other advisors.

We then evaluate the above 4 components independently:

$$P(o_{u,a} | t_{u,a} = t, o_{u,-a}, \varepsilon) = \frac{n(o_u^1|t) + \varepsilon}{n(o_u^0|t) + n(o_u^1|t) + 2\varepsilon}$$

where $n(o_u^1|t)$ and $n(o_u^0|t)$ are the numbers of times that user u shows and hides the trust t to advisors, respectively. $n(o_u^1|t)$ and $n(o_u^0|t)$ must exclude the counts between u and a because of the condition on $o_{u,-a}$.

$$P(t_{u,a} = t | t_{u,-a}, \alpha, \psi) = \left[\frac{n(t_a^t) + \alpha}{n(t_a^0) + n(t_a^1) + 2\alpha}\right] \left[\frac{n(t_u^t) + \psi}{n(t_u^0) + n(t_u^1) + 2\psi}\right]$$

where $n(t_u^t)$ is the number of links with trust value t from user u, and $n(t_a^t)$ is the number of links with value t to advisor a. Similarly, due to $t_{u,-a}$, $n(t_u^t)$ and $n(t_a^t)$ must exclude the counts between u and a.

$$
\begin{aligned}
&P(R_{U_{-a},a}|t_{u,a}=t,R_{U_{-a},-a},\theta)\\
=&\Bigg[\Gamma\left(\sum_{r=0}^{m} m.n(R_a^r|t)+2\theta\right)\Gamma\left(\sum_{r=0}^{m} r\left[n(R_a^r|t)+n(R_{U_{-a},a}^r|t)\right]+\theta\right)\\
&\Gamma\left(\sum_{r=0}^{m}(m-r).\left[n(R_a^r|t)+n(R_{U_{-a},a}^r|t)\right]+\theta\right)\Bigg]\\
&\Bigg[\Gamma\left(\sum_{r=0}^{m} r.n(R_a^r|t)+\theta\right)\Gamma\left(\sum_{r=0}^{m}(m-r).n(R_a^r|t)+\theta\right)\\
&\left(m\sum_{r=0}^{m}\left[n(R_a^r|t)+n(R_{U_{-a},a}^r|t)\right]+2\theta\right)\Bigg]^{-1}
\end{aligned}
$$

where $n(R_a^r|t)$ refers to the number of rating difference r between a with all the other users if a has been given trust t. $n(R_{U_{-a},a}|t)$ denotes the number of rating difference r between a and the average of all the other ratings with regard to commonly rated entities with u given that u's trust on a is t.

$$
\begin{aligned}
&P(r_{u,a}|t_{u,a}=t,r_{u,-a},\varphi,\beta)=\\
&\Bigg[\Gamma\left(\sum_{r=0}^{m} m.n(r_a^r|t)+2\varphi\right)\Gamma\left(\sum_{r=0}^{m} r\left[n(r_a^r|t)+n(r_{u,a}^r|t)\right]+\varphi\right)\\
&\Gamma\left(\sum_{r=0}^{m}(m-r)\left[n(r_a^r|t)+n(r_{u,a}^r|t)\right]+\varphi\right)\Bigg]\Bigg[\Gamma\left(\sum_{r=0}^{m} r.n(r_a^r|t)+\varphi\right)\\
&\Gamma\left(\sum_{r=0}^{m}(m-r)n(r_a^r|t)+\varphi\right)\left(m\sum_{r=0}^{m}\left[n(r_a^r|t)+n(r_{u,a}^r|t)\right]+2\varphi\right)\Bigg]^{-1}\\
&\Bigg[\Gamma\left(\sum_{r=0}^{m} m.n(r_{u,a}^r|t)+2\beta\right)\Gamma\left(\sum_{r=0}^{m} 2r\left[n(r_{u,a}^r|t)\right]+\beta\right)\\
&\Gamma\left(\sum_{r=0}^{m} 2(m-r).\left[n(r_{u,a}^r|t)\right]+\beta\right)\Bigg]\Bigg[\Gamma\left(\sum_{r=0}^{m} r.n(r_{u,a}^r|t)+\beta\right)\\
&\Gamma\left(\sum_{r=0}^{m}(m-r).n(r_{u,a}^r|t)+\beta\right)\left(2m\sum_{r=0}^{m} n(r_{u,a}^r|t)+2\beta\right)\Bigg]^{-1}
\end{aligned}
$$

where $n(r_{u,a}^r|t)$ and $n(r_a^r|t)$ denote the number of rating difference r between a and u given trust value t, and the number of rating difference r between a and other users given that trust value to a is t, respectively.

When trust links are not observable, as shown in Figure 2(b). The corresponding sampling process is:

$$
\begin{aligned}
&P(t_{u,a}=t|r,R,\alpha,\varphi,\theta,\beta,\psi)\propto P(t_{u,a}=t|\alpha,\psi)\\
&P(r_{u,a}|t_{u,a}=t,r_{u,-a},\varphi,\beta)P(R_{U_{-a},a}|t_{u,a}=t,R_{U_{-a},-a},\theta)
\end{aligned}
$$

The inference process is similar to the Gibbs sampling process where the trust links are partially observable. The only difference is that we need to conduct sampling for each trust link between users and advisors.

Parameter Estimation

After the inference on $t_{u,a}$, we can update posterior distributions of y, c, e, i, b and s, as follows:[2]

$$
\begin{aligned}
&P(y_u|t_u,\psi,\alpha)\sim Beta\left(\psi+n(t_u^1),\psi+n(t_u^0)\right)\\
&P(c_a|t_a,\psi,\alpha)\sim Beta\left(\alpha+n(t_a^1),\alpha+n(t_a^0)\right)\\
&P(e_{u|t}|t_u,o_u,\varepsilon)\sim Beta\left(\varepsilon+n(o_u^1|t),\varepsilon+n(o_u^0|t)\right)\\
&P(s_{u,a|t}|t_a,r_a,b_{a|t},\beta)\sim\\
&\qquad Beta\left(\sum_{r=0}^{m} r\,n(r_{u,a}^r|t)+\beta,\sum_{r=0}^{m}(m-r)\,n(r_{u,a}^r|t)+\beta\right)\\
&P(i_{a|t}|t_a,R_a,\theta)\sim Beta\Bigg(\sum_{r=0}^{m}\left[r(1-t)n(R_a^r|t)+t(m-r)n(R_a^r|t)\right]\\
&\qquad+\theta,\sum_{r=0}^{m}\left[rtn(R_a^r|t)+(1-t)(m-r)n(R_a^r|t)\right]+\theta\Bigg)
\end{aligned}
$$

[2]Due to space limitation, we omit detailed derivation process.

$$
\begin{aligned}
&P(b_{a|t}|t_a,r_a,s_a,\varphi)\sim Beta\Bigg(\sum_{r=0}^{m}\left[r(1-t)n(r_a^r|t)+t(m-r)n(r_a^r|t)\right]\\
&\qquad+\varphi,\sum_{r=0}^{m}\left[rtn(r_a^r|t)+(1-t)(m-r)n(r_a^r|t)\right]+\varphi\Bigg)
\end{aligned}
$$

3.3 Trust Prediction

After learning the parameters, we can evaluate the trust between user u and advisor a, $t_{u,a}$, using the Markov blanket [Pearl, 1988] of node t in Figure 2. The Markov blanket of a node contains all the variables that shield the node from the rest of the network: its parent, its children and its children's parents. This determines that, for our scenario, the Markov blanket of node t becomes the only knowledge needed to predict the behavior of the node t. Thus, we identify the probability of $t_{i,j}=t$ as follows:

$$
\begin{aligned}
&P(t_{u,a}=t|o_{u,a},r_{u,a},R_{U_{-a},a},c_a,b_a,i_a,s_{u,a},y_u,e_u)\\
&\propto P(t_{u,a}=t|y_u,c_a)P(R_{U_{-a},a}|t_{u,a}=t,i_a)\\
&\quad P(o_{u,a}|t_{u,a}=t,e_u)P(r_{u,a}|t_{u,a}=t,b_a,s_{u,a})
\end{aligned}
$$

We can use the probability value ($t_{u,a}=1$) as the trust value if we expect to obtain continues trust values ranged in $[0,1]$.

4 Experiments

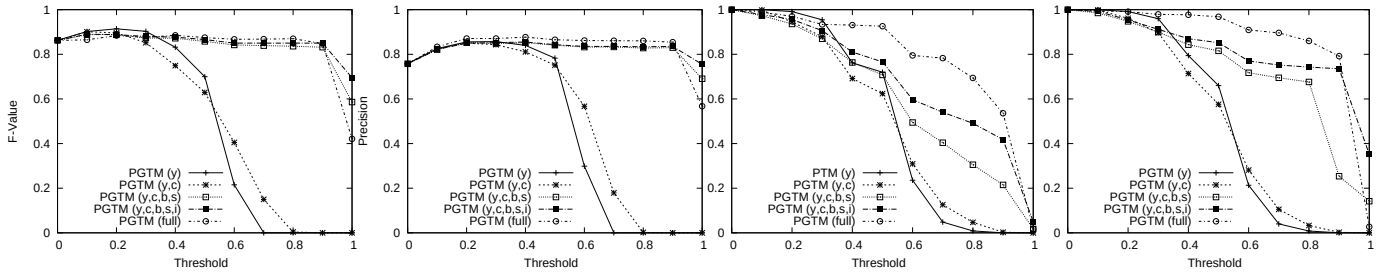
In this section, we carry out experiments to evaluate the performance of our probabilistic graphical trust model (PGTM) on predicting advisors' trustworthiness perceived by users, and conduct comparisons with some competing approaches.

4.1 Benchmark Approaches

We compare our approach with two state-of-the-art models, including BLADE [Regan *et al.*, 2006] and Prob-Cog [Noorian *et al.*, 2011] detailed in Section 2. For BLADE implementation, we treat the average ratings of entities as the attribute modeled on the three datasets. For the Prob-Cog, we tune the parameters so that the model achieves its best performance. Specifically, in the first layer, an advisor is considered as dishonest when its rating difference with a user is larger than threshold μ ($\mu\in[0.5,0.8]$). In the second layer, the trustworthiness of an advisor is further adjusted by her tendency (positive or negative) ($\beta+\epsilon\leq\mu$). We also show the performance of a naive baseline approach where a user judges an advisor's trustworthiness based on commonly rated entities. If the rating difference on a same entity is smaller than a predefined threshold, the user's experience with the advisor on the entity is positive, otherwise negative. Positive and negative experiences are aggregated to compute the advisor's trustworthiness using the Beta function.

4.2 Data Description

Each of our dataset consists of two files. One file stores the (dis)trust links with 3-tuples $(user,user,(dis)trust)$ which serve as ground-truth, and the other file stores users' ratings of entities with 3-tuples $(user,entity,rating)$. Notice that the links are directed: user a (dis)trusts user b does not imply that b also (dis)trusts a. The goal of all models is to predict the ground-truth (dis)trust links based on the commonly



(a) Epinions (F-Value) **(b) Epinions (Precision)** **(c) Flixster (Precision)** **(d) FilmTrust (Precision)**

Figure 3: Model Effectiveness by Varying the Trust Threshold

rated entities between users. To obtain our data, we process three available datasets, including FilmTrust (`trust.mindswap.org/FilmTrust`), Flixster.com and Epinions.com. Specifically, we use the original FilmTrust dataset, and randomly sample a small portion of original Flixster dataset. For Epinions dataset, the (dis)trust links between users might be due to their direct rating interactions since the entity (i.e. article) is created by users who in turn could rate others' entities. We thus exclude all (dis)trust links in which users have rated some entities created by advisors. The statistical information is summarized in Table 1.

Datasets	Epinions	Flixster	FilmTrust
Trust value	0,1	1	1
Rating scale	1-5	1-10	1-8
Users	999	617	874
Entities	545,499	4,683	1,957
Ratings	2,089,872	18,436	18,662
Trust links	753	453	1,437
Distrust links	240	-	-
Avg. commonly rated entities	71.4	4.71	8.00

Table 1: Statistical Information about the Three Datasets

4.3 Evaluation Metrics

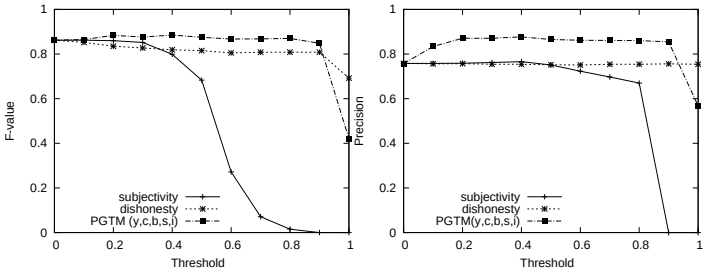
To measure the performance of models, we use the commonly used metrics, including *precision*, *recall*, *f-value*, and *MAE*. Precision=$\frac{t_p}{t_p+f_p}$, recall=$\frac{t_p}{t_p+f_n}$, and f-value=$\frac{2\times\text{precision}\times\text{recall}}{\text{precision}+\text{recall}}$, where t_p, f_p and f_n is the number of correctly predicted trust links, incorrectly predicted trust links, and incorrectly predicted distrust links respectively. MAE (mean absolute error) refers to the average of the difference between the predicted trust value of each user-advisor pair and the ground-truth trust value (0 or 1). Since there are merely trust links but no distrust links on FilmTrust and Flixster datasets, only *precision* and *MAE* are used for these two datasets.

4.4 Results and Discussion

In this section, we first check the effectiveness of latent variables in our model. Then, we present the performance of our model and three benchmark approaches on the three datasets. We further examine these approaches in detail by varying the threshold for trust prediction. If a predicted trust value of advisor a from user u is larger than the threshold, u trusts a.

Model Effectiveness

In Figure 3, we analyze the effectiveness of each latent variable in our model by showing the performance of the



(a) **(b)**

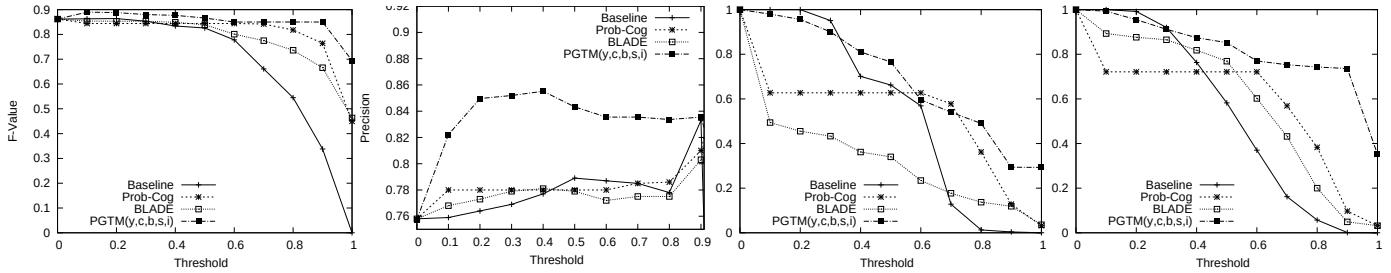
Figure 4: Model Effectiveness (Subjectivity vs. Dishonesty)

model with only *propensity* latent variable (PGTM(y)), with *propensity* and *competence* latent variables (PGTM(y, c)), with *propensity*, *competence*, *benevolence* and *subjectivity difference* latent variables (PGTM(y, c, b, s)), with all but *expressiveness* latent variable (PGTM(y, c, b, s, i)) and the complete model (PGTM($full$)) which includes all latent variables. As can be seen, PGTM(y, c) performs better than PGTM(y) when trust threshold is larger than 0.5. This might be due to the fact that PGTM(y, c) rather than PGTM(y) could enable the risk-aversion users (with higher trust threshold) to adjust their propensity to (dis)trust according to advisors' competence through interactions. Next, the superiority of PGTM(y, c, b, s) over PGTM(y, c) becomes more salient when the number of commonly rated entities between users and advisors increases (**Epinions**>**FilmTrust**>**Flixster** as shown in Table 1). This is in accordance with the structure of our probabilistic trust model where *subjectivity difference* between users and advisors and the *benevolence* of advisors are directly connected with the rating difference between users and advisors. With more commonly rated items, we could model rating difference more accurately. Moreover, PGTM(y, c, b, s, i) could outperform PGTM(y, c, b, s) saliently when the number of commonly rated items between a user and an advisor is limited. This is because that the additional variable *integrity* is mainly modeled by the rating difference between the advisor and all other users and thus is less sensitive to the number of commonly rated items. Finally, the complete model PGTM($full$) performs best because instead of using the ratings, we use partially observable trust links to model the *expressiveness* latent variable. Here, 20% of the trust links are observable.

We also investigate the effects of *subjectivity difference* between users and advisors and *dishonesty* of advisors in our model. The result on Epinions dataset is illustrated in Figure 4, where *subjectivity* denotes the model with only *propen-*

Approach \ Dataset	Epinions				Flixster		FilmTrust	
	Precision	**Recall**	**F-value**	**MAE**	**Precision**	**MAE**	**Precision**	**MAE**
PGTM (y, c, b, s, i)	**0.843**	0.896	**0.866**	**0.077**	**0.766**	**0.296**	**0.848**	**0.161**
Baseline	0.789	0.866	0.826	0.374	0.662	0.452	0.582	0.477
BLADE	0.779	0.910	0.839	0.312	0.340	0.665	0.769	0.394
Prob-Cog	0.780	**0.920**	0.844	0.281	0.627	0.468	0.721	0.422

Table 2: Performance Comparison on the FilmTrust, Epinions and Flixster Datasets



(a) Epinions (F-Value) **(b) Epinions (Precision)** **(c) Flixster (Precision)** **(d) FilmTrust (Precision)**

Figure 5: Performance Comparison by Varying the Trust Threshold

sity and *subjectivity difference* latent variables, while *dishonesty* denotes the model with *propensity* and *competence*, *benevolence*, and *integrity* latent variables. It can be observed that PGTM(y, c, b, s, i) outperforms *subjectivity* model and *dishonesty* model, demonstrating that the subjectivity difference factor which is paid less attention before, in addition to the dishonesty factor, improve the performance of our model. We also find that the dishonesty plays a more important role than subjectivity difference. Specifically, the omission of the dishonesty related latent variables (i.e. competence, benevolence and integrity) will lead to a more salient drop than the omission of subjectivity-related latent variable (i.e., subjectivity difference) when the trust threshold is larger than 0.5.

Model Comparison

Table 2 and Figure 5 show the performance comparisons between our approach (PGTM(y, c, b, s, i)) and other approaches on three datasets. In order to make fair comparisons, we assume that all trust links are not observable in our model (i.e. we adopt the model in Figure 2 (b)). First of all, although our approach is a little more time-consuming than other approaches, we find that the running time is acceptable. Specifically, it takes $0.312s$ ($0.061s$ for Prob-Cog and $0.218s$ for BLADE) to run each iteration of our model on Epinions dataset. Second, as shown in Table 2, our model (with trust threshold 0.5) achieves much better performance than other approaches in terms of all metrics on all three datasets (except *recall* on Epinions). The performance of BLADE and Prob-Cog are better than Baseline on the Epinions and FilmTrust, but worse than that on the Flixster. This is mainly because there are fewer commonly rated entities in Flixster dataset than in other two datasets (see Table 1). Without enough commonly rated items, Prob-Cog may mistakenly treat subjective users as dishonest ones, and thus filter them in the first layer, while BLADE cannot model advisors' evaluation functions on attribute accurately. On the other hand, our model is not so sensitive to the number of commonly rated entities, because only *subjectivity difference* between a user and an advisor is measured with regard to the commonly rated entities. *Benevolence*, *competence* and *integrity* are measured based on the advisor's past experience with all other users, and *propensity* is measured according to all the past experiences of the user.

Figure 5 presents the performance of different approaches by varying the trust threshold. It shows that in general our model outperforms the other approaches. It consistently achieves high precision on the three datasets, demonstrating its effectiveness on modeling the trustworthiness of advisors. It also implies the ability of our model on inferring the (unobserved) (dis)trust links. This is especially important in the current online communities, where users are reluctant to explicitly identify their relationship with other users.

5 Conclusions and Future Work

We proposed a novel probabilistic graphical trust model, separately considering *dishonesty* of advisors and their *subjectivity difference* between users, to model the trustworthiness of advisors. Specifically, our model involves three types of latent variables: 1) the dishonesty related variables: *benevolence*, *integrity* and *competence* of advisors; 2) *trust propensity* of users; and 3) *subjectivity difference* between users and advisors. We compared our model with a baseline approach, and two state-of-the-art approaches including BLADE and Prob-Cog. Experimental results indicated that the latent variables in our model are both theoretically reasonable and computationally effective, and dishonesty and subjectivity difference are successfully distinguished. Besides, we demonstrated that our model can more accurately model advisor trustworthiness without using the partially observable trust links. Our approach also mitigates the research gap between computational trust in Computer Science and psychological and behavioral trust in Social Science. For future work, we will extend the current model to address other scenarios (e.g., multi-nominal degrees of trust other than binary case).

Acknowledgement

The work is supported by the MoE AcRF Tier 2 Grant M4020110.020 and the Institute for Media Innovation at Nanyang Technological University.

References

[Casella and George, 1992] George Casella and Edward I. George. Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.

[Chua and Lim, 2010] Freddy Chong Tat Chua and Ee-Peng Lim. Trust network inference for online rating data using generative models. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 889–898, Washington, DC, USA, 2010. ACM.

[Fang *et al.*, 2012] Hui Fang, Jie Zhang, Murat Sensoy, and Nadia Magnenat-Thalmann. SARC: subjectivity alignment for reputation computation. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, pages 1365–1366, 2012.

[Jordan *et al.*, 1999] Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. An introducation to variational methods for graphical models. *Machine Learning*, 37:183–233, 1999.

[Liu *et al.*, 2011] Siyuan Liu, Jie Zhang, Chunyan Miao, Yin-Leng Theng, and Alex C Kot. iclub: An integrated clustering-based approach to improve the robustness of reputation systems. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*, pages 1151–1152. International Foundation for Autonomous Agents and Multiagent Systems, 2011.

[Liu, 1994] Jun S. Liu. The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966, 1994.

[Mayer *et al.*, 1995] Roger C. Mayer, James H. Davis, and F. David Schoorman. An integrative model of organizational trust. *The Academy of Management Review*, 20(3):709–734, July 1995.

[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.

[Noorian *et al.*, 2011] Zeinab Noorian, Steve Marsh, and Michael Fleming. Multi-layered cognitive filtering by behavioural modelling. In *Proceedings of the 10th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 871–878, 2011.

[Pearl, 1988] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo CA, 1988.

[Regan *et al.*, 2006] Kevin Regan, Pascal Poupart, and Robin Cohen. Bayesian reputation modeling in E-marketplaces sensitive to subjectivity, deception and change. In *Proceedings of the 21st National Conference on Artificial Intelligence*, pages 1206–1212, Boston, Massachusetts, USA, 2006. AAAI Press.

[Teacy *et al.*, 2006] W. T. L. Teacy, J. Patel, N. R. Jennings, and M. Luck. TRAVOS: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12(2):183–198, 2006.

[Zhang and Cohen, 2007] Jie Zhang and Robin Cohen. A comprehensive approach for sharing semantic web trust ratings. *Computational Intelligence*, 23(3):302–319, 2007.