Coping with subjectivity and dishonesty in opinion evaluation by exploiting social factors

Fang, Hui

2014

Fang, H. (2014). Coping with subjectivity and dishonesty in opinion evaluation by exploiting social factors. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/62116>

<https://doi.org/10.32657/10356/62116>

**COPING WITH SUBJECTIVITY AND DISHONESTY
IN OPINION EVALUATION BY EXPLOITING
SOCIAL FACTORS**

FANG HUI

School of Computer Engineering

A thesis submitted to the Nanyang Technological University
in fulfilment of the requirements for the degree of
Doctor of Philosophy

2014

Acknowledgments

This dissertation would not have been possible without the help and support of lots of individuals and institutions, to only some of whom are possible to give particular mention here.

Above all, I would like to thank my principle supervisor Dr. Jie Zhang for his constant mentorship and full support in the past four years. He has taught me, both consciously and unconsciously, how good research is done. He has provided me extensive research guidance, and helped me to develop the academic skills in paper writing, oral presentation, and critical thinking. All these skills have prepared me well to carry out independent research. The good advice, constant support and kindness of my second supervisor, Dr. Nadia Magnenant-Thalmann, have been invaluable on my academic career, for which I am extremely grateful.

Special thanks go to my Qualification Examination (QE) committee members Dr. Chunyan Miao, Dr. Anwitaman Datta and Dr. Yunke Chang, for their critical reading and valuable feedback on my QE report, which helped me better schedule my research plan in the last two years of my Ph.D. study.

I would also like to thank many other professors and research staffs who have supported me and provided me valuable suggestions for my work, especially, Dr. Murat Şensoy, Dr. Qinghua Zhu, Dr. Kate Larson, Dr. Xin Liu, Ms. Feyza Merve Hafizoglu, Dr. Zeinab Noorian, and Mr. Lizi Zhang. Particular thanks are given to Dr. Murat Şensoy from Ozyegin University, Turkey for his guidance and unsurpassed suggestions on my research, and warm friendship on my personal life.

I am fortune to have a multitude of great friends and colleagues who have supported me both research wise and non-research wise in Nanyang Technological University, particularly, Ms. Yuan Liu, Mr. Siwei Jiang, Mr. Guibing Guo, Ms. Athirai A. Irissappane,

Ms. Huanhuan Zhang, Mr. Chang Xu, Ms. Peng Cheng, Ms. Lu Dong, Mr. Quan Yuan, Ms. Yuzhe Zhang, and Mr. Hongyuan Zhu. Without their advice and friendship, my Ph.D. journey would have been a less enjoyable one.

I gratefully acknowledge the funding sources that made my Ph.D. study possible. I was funded by the research scholarship from Institute for Media Innovation (IMI) for my four-year graduate study, and supported by the IMI and SCE with generous travel grants to attend academic conferences for presenting and discussing the works in the dissertation. I also thank the staffs coming from the IMI and SCE, especially Ms. Woo Elsie, Ms. Poh Yian Lim, Ms. May Thu Aung, Mr. Qui C.T. Tran, Ms. Len Ah Chan and Mr. Chng Hee Siah, for their technical support and administrative assistance.

Lastly and most importantly, I would like to thank all of my family members who have given me a lifetime of love and support, and have always been a source of encouragement for me. I give my special thanks to my husband Yang. His constant patience, encouragement, and care sustain me.

Summary

In large and open environments, users may often encounter entities (e.g. people, products and information) which they have no previous experience with or prior knowledge of. In this case, they usually rely on the experience or knowledge of other users (referred as advisors) in the form of opinions (e.g. reviews and ratings), to choose which entities to interact with. However, in these environments, users can freely express their opinions, and the quality of opinions may then vary. Of all the possible reasons that result in diversity of opinion quality, the following two are the most dominant: (1) users are subjectively different, which thus leads to discrepancy of users' opinions towards same entities; and (2) some users might be dishonest and lie about their experience with entities. However, some existing approaches may only consider either dishonesty or subjectivity difference, while others cannot accurately distinguish these two different aspects.

On the other hand, subjectivity difference and dishonesty, regarded as either the extrinsic or intrinsic characteristics of users (human-beings), are more related to the behavioral and psychological perspectives of users other than merely the statistical probability of data points that prevalently adopted by Computer Science. In other words, it is expected to involve understanding users' motivation and needs of providing opinions when evaluating them. It is thus worthwhile to incorporate social factors when addressing these two aspects for opinion evaluation.

With the aforementioned two issues in mind, we aim to solve the subjectivity and dishonesty problems for opinion evaluation by also taking social factors into consideration in this dissertation. In particular, this dissertation is comprised of four studies.

First, we propose a novel trust model stemmed from diffusion theory in Social Science (called DiffTrust), to evaluate the opinions of advisors by modeling their trustworthiness. 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. On the other side, diffusion theory 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. 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. DiffTrust can capture the dynamics of trust, and its dependency on other users and the environment. However, it cannot accurately distinguish subjectivity and dishonesty.

Second, we design a subjectivity alignment approach for reputation computation (SARC) when aggregating numerical ratings (opinions) provided by users towards the same entities. Subjectivity difference between users may come from two sources by analyzing the scenario of a user providing a rating towards an entity, from both psychological and behavioral perspectives: (1) intra-attribute subjectivity, the subjectivity in evaluating the same attribute; and (2) extra-attribute subjectivity, the subjectivity in evaluating different attributes. We learn these two kinds of subjectivity for each user by applying Bayesian learning and regression analysis on the basis of each user's past experience, respectively. Ratings provided by one user can thus be aligned for another user according to the two users' subjectivity. Although SARC only considers the subjectivity difference aspect, it is not much affected by dishonest users, as validated in our experiments.

Finally, we also propose two approaches by explicitly distinguishing subjectivity and dishonesty. In the first approach, we present a novel probabilistic graphical trust model (PGTM) to separately consider these two aspects. We model the factors of advisors' intrinsic nature (dishonesty, i.e. benevolence, integrity and competence from Social Science), 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. PGTM ignores the fact that dishonesty and subjectivity overlap with each

other to certain extent. Therefore, we further propose a two-layered clustering approach (SubGroup) by modeling each advisor as part of groups. Specifically, on the basis of indicative features extracted from users' rating behavior to distinguish subjectivity and dishonesty, in the first layer, each user clusters her advisors into different subjectivity groups and dishonest types, with respect to their rating behavior. In the second layer, each advisor is assigned to two groups with respective membership degrees. An alignment approach is designed to help each user align advisors' ratings to the ones of her own.

In summary, the main technical contributions of this dissertation are three-folds: (1) DiffTrust, stemmed from the diffusion theory in Social Science, mainly deals with advisors' dishonesty; (2) SARC addresses two kinds of subjectivity difference between users; and (3) PGTm and SubGroup approaches can well distinguish and cope with dishonesty and subjectivity difference for opinion evaluation. As a new attempt to consider social factors in trust assessment, our approaches contribute to bridging the research gap between computational trust in Computer Science and psychological and behavioral trust in Social Science. Further, we hope that these in-depth studies induce more attention towards this important interdisciplinary research direction .

Contents

Acknowledgments	ii
Summary	iv
Contents	vii
List of Figures	xi
List of Tables	xiv
1 Introduction	1
1.1 Research Problems	3
1.2 Research Scope and Solutions	5
1.2.1 Social Theory Adaption for Trust Modeling	5
1.2.2 Social Factors Concerning Subjectivity	6
1.2.3 Social Factors Concerning Subjectivity and Dishonesty	7
1.3 Research Contributions	8
1.4 Dissertation Outline	10
2 Literature Review	12
2.1 Online Opinions	12
2.1.1 Social Influence on Decision-making	13
2.1.2 Quality Problems	15
2.2 Studies Related to Opinion Evaluation	16
2.2.1 Reputation Mechanisms	17
2.2.2 Computational Trust Models in MAS	18
2.2.3 Recommender Systems	25
2.3 Summary	28

3	DiffTrust: A Trust Model Stemmed from Diffusion Theory	30
3.1	The DiffTrust Model	31
3.1.1	Trust and the Diffusion Theory	31
3.1.2	The Trust Model	33
3.1.3	Trust Computation	35
3.2	Experiments	40
3.2.1	Data Acquisition	41
3.2.2	Evaluation Metrics	43
3.2.3	Results and Discussion	44
3.3	Summary	49
4	SARC: Subjectivity Alignment for Reputation Computation	51
4.1	Overview of the SARC Approach	53
4.2	Technical Details of the SARC Approach	56
4.2.1	Intra-attribute Subjectivity Alignment	56
4.2.2	Extra-attribute Subjectivity Alignment	59
4.3	Experiments	60
4.3.1	Experimental Environment	61
4.3.2	Experimental Parameters	62
4.3.3	Experimental Results	64
4.3.4	Discussions on Experimental Results	70
4.4	Summary	71
5	PGTM: Probabilistic Graphical Trust Model	73
5.1	The Probabilistic Graphical Trust Model	74
5.1.1	Parameters and Generative Process	75
5.1.2	Model Inference and Parameter Estimation	79
5.1.3	Trust Prediction	83
5.2	Experiments	83

5.2.1	Data Description	83
5.2.2	Benchmark Approaches	84
5.2.3	Evaluation Metrics	85
5.2.4	Results and Discussion	85
5.3	Summary	92
6	SubGroup: Learning From Users' Rating Behavior	93
6.1	Subjectivity and Dishonesty	95
6.1.1	Subjectivity	95
6.1.2	Dishonesty	96
6.2	The SubGroup Approach	96
6.2.1	Procedural Framework	97
6.2.2	Feature Identification	98
6.2.3	Cluster Analysis	101
6.2.4	Group Alignment	105
6.3	Experiments	106
6.3.1	Simulated E-marketplace	106
6.3.2	Real Environments	110
6.4	Summary	113
7	Discussion	114
7.1	Model Comparison	114
7.1.1	Effectiveness Comparison	114
7.1.2	Robustness Comparison	117
7.2	Adapting Trust Factors into Recommender Systems	119
7.2.1	The (Dis)Trust Framework	121
7.2.2	Trust and Distrust Prediction	125
7.2.3	Evaluation	130
7.2.4	Results and Analysis	133

7.3	Summary	143
8	Conclusions and Future Work	145
8.1	Concluding Remarks	145
8.2	Future Work	149
8.2.1	Extending the Current Trust Models	149
8.2.2	A Testbed for Evaluating Trust Models	150
8.2.3	Checking the Robustness of Trust Models in Resisting Attacks	151
8.2.4	Merging Review Evaluation with Rating Evaluation	152
8.2.5	Incorporating Other Suitable Social Theories	153
	References	154

List of Figures

1.1	The Relationship in the Online Community	2
3.1	The DiffTrust Model for Evaluating the Trustworthiness of Advisors	34
3.2	Performance Comparison on the eBay Dataset by Varying the Time Slots (MAE)	45
3.3	Performance Comparison on the eBay Dataset by Varying the Time Slots (MCC)	46
3.4	Performance Comparison on the eBay Dataset by Varying the Trust Threshold (MCC)	47
3.5	Performance Comparison on the FilmTrust by Varying the Trust Threshold	48
3.6	Performance Comparison on the Epinions by Varying the Trust Threshold	49
3.7	Performance Comparison on the Flixster by Varying the Trust Threshold	50
4.1	Overview of the SARC Approach	55
4.2	A Naïve Bayesian Network Model for Agent a_i of Buyer b_i to Align b_k 's Rating r^{b_k}	57
4.3	Performance Comparison in the Basic Environment Where $R_{liar} = 0, P_{seller} = 0, P_{Buyer} = 0$	65
4.4	Performance When Varying the Ratio of Objective Attributes	65
4.5	Performance When Varying the Granularity of Rating Scale	66
4.6	Performance When Varying the Number of Detailed Reviews	66
4.7	Varying Ratio of Shared Interactions	67
4.8	Performance When Varying Ratio of Lying Buyers	68
4.9	Performance When Varying Probability of Sellers' Changing Behavior	69
4.10	Performance When Varying Probability of Buyers' Changing Subjectivity	70

5.1	The Conceptual Framework of Trust	75
5.2	Graphical Model: (a) with (b) without Trust Links	78
5.3	Model Effectiveness on the Epinions Dataset by Varying the Trust Threshold (F-Value)	86
5.4	Model Effectiveness on the Epinions Dataset by Varying the Trust Threshold (Precision)	86
5.5	Model Effectiveness on the Flixster Dataset by Varying the Trust Threshold	87
5.6	Model Effectiveness on the FilmTrust Dataset by Varying the Trust Threshold	87
5.7	Model Effectiveness (Subjectivity vs. Dishonesty) (F-Value)	88
5.8	Model Effectiveness (Subjectivity vs. Dishonesty) (Precision)	89
5.9	Performance Comparison on the Epinions Dataset by Varying the Trust Threshold (F-Value)	90
5.10	Performance Comparison on the Epinions Dataset by Varying the Trust Threshold (Precision)	90
5.11	Performance Comparison on the Flixster Dataset by Varying the Trust Threshold	91
5.12	Performance Comparison on the FilmTrust Dataset by Varying the Trust Threshold	91
6.1	Procedural Design of the Framework	98
6.2	Performance Change of Our SubGroup Model by Varying σ_1	108
6.3	Performance Change of Our SubGroup Model by Varying σ_2	108
6.4	Performance Change of Our SubGroup Model by Varying Number of Features ($\sigma_1 = 0.12, \sigma_2 = 0.30$)	109
6.5	Performance Change of Our SubGroup Model with Different Component	109
6.6	Performance Comparison as the Change of Iterations	110
6.7	Performance Comparison When Varying the Ratio of Liars	111
7.1	The Proposed (Dis)trust Framework	122
7.2	The Comparison of Refined Trust Links with the Original Ones	135

7.3	Performance Comparison of TidalTrust Method by Considering Interpersonal and Impersonal Aspects	138
7.4	Performance Comparison of TidalTrust Method in All View On Epinions1	138
7.5	Performance Comparison of TidalTrust Method in Different Views on Epinions1	139
7.6	Performance Comparison of TidalTrust Method in All View	141
7.7	Performance by Varying the Size of Predicted Trust Network on FilmTrust	141
7.8	Performance Comparison of TidalTrust between “without” and “with” Distrust	142

List of Tables

3.1	Summary of Notations in DiffTrust	33
3.2	Statistical Information about the Four Real Datasets	42
3.3	Performance Comparison on the eBay Dataset	44
3.4	Performance Comparison on the FilmTrust, Epinions and Flixster Datasets	48
4.1	Summary of Notations in SARC	54
4.2	Product Attributes and Value Ranges	62
5.1	Summary of Notations in PGMT	76
5.2	Statistical Information about the Three Datasets	84
5.3	Performance Comparison on the FilmTrust, Epinions and Flixster Datasets	89
6.1	The Statistics of the Three Datasets	112
6.2	Performance Comparison of Different Approaches	112
7.1	Comparison of the Four Models	115
7.2	Model Comparison in Resisting Attacks	118
7.3	The Statistics of the Four Datasets	131
7.4	The Coefficients of Trust Aspects	133
7.5	Performance Comparison Based on Refined Trust Using Epinions ¹	136
7.6	Performance Comparison Based on Different Trust Aspects Using Epinions ²	137
7.7	Performance Comparison of Tidaltrust Based on Interpersonal Aspects on Flixster	140

CHAPTER 1

Introduction

The Internet has become an inseparable part of our daily life nowadays. According to the Internet World Stats¹, the number of Internet users worldwide has reached 2.41 billion by the end of June 2012, accounting for 34.3 percent of the global population. Users have different channels to connect to the Internet, including dial-up connections, broadband connections, and mobile connections (Pearce and Rice, 2013). The increasing availability of the Internet attracts more and more people to go online, be involved in a wide range of online activities, and indulge in various kinds of online communities, such as forums, blogs, games, newsgroups and social networks, where different people have been brought together to interact with one another (Zhang et al., 2011). They share information, express opinions and also make friends. There are a variety of popular online communities in the world, such as expert sites like Askme (askmecorp.com) and Advocate (advogato.org), products review sites like Epinions (epinion.com), Ciao (ciao.com) and Flixster (flixster.com), discussion forums like Slashdot (slashdot.org) and Douban (douban.com), social networks like Facebook (facebook.com), Twitter (twitter.com), Flickr (flickr.com) and Weibo (weibo.com), and e-commerce systems like eBay (eBay.com), Amazon (amazon.com) and Taobao (taobao.com).

In large and open online communities, users have a low overhead of joining and leaving the systems, and also can easily change their virtual identities. Due to the dynamic and uncertain characteristics of these environments, users may often encounter entities (e.g. hotels, books, movies, political candidates, news, comments, and stories)

¹<http://www.internetworldstats.com/stats.htm>

Figure 1.1: The Relationship in the Online Community

which they have no previous experience with or prior knowledge of. Some users might thus suffer from the potential risks of monetary loss or security issues when interacting with malicious entities. In this case, users usually rely on the experience or knowledge of other users, to choose which entities to interact with. Other users who have previously interacted with the entities could report and describe their experience usually in the form of numerical ratings or textual reviews (i.e. opinions), and thus are referred as *advisors*² in our research scenario. Users depend on online opinions contributed by these advisors to make wise decisions, on which entities are worth their time and money (Muchnik et al., 2013). In other words, users are interested in whether they can successfully harness the wisdom of crowds (advisors) and thus evaluate the unknown target entities without bias. The basic relationship among user, advisor and entity in an online community is presented in Figure 1.1.

However, in online environments, advisors can freely express their opinions with limited administration, and the quality of opinions may then vary (as indicated in Figure 1.1). On the one hand, some advisors may be dishonest and lie about their experience (Whitby et al., 2004; Zhang and Cohen, 2008; Yang et al., 2009). For example, to demote an

²Advisors are also users, thus in this dissertation, we use these two terms interchangeably.

entity, a dishonest advisor may intentionally provide a negative rating to the entity while the real experience is satisfactory. On the other hand, people are subjectively different due to their intrinsic characteristics or past experience. Advisors provide true opinions based on their experience, which might be *unintentionally* misleading for some users due to the inherent subjectivity of users' opinions (Regan et al., 2006; Ghose and Ipeirotis, 2007; Teacy et al., 2012; Fang et al., 2012b). For instance, a satisfactory experience of an advisor may turn to be an unsatisfactory one for a user because they value similar experience differently. Therefore, it is important for users to evaluate the quality of advisors' opinions in order to determine how much to rely on: *do they produce useful, unbiased and truthful information about the quality of the entities being evaluated?*

1.1 Research Problems

As aforementioned, there are mainly two reasons leading to the diversified quality of advisors' opinions with regard to users: (1) advisors might be dishonest and untruthfully report their experience; and (2) there exists the subjectivity difference between users and advisors. It thus becomes crucially important to tackle the two issues in order to effectively address the opinion quality problem. It should be noted that in this dissertation, online opinions are mainly referred to as *ratings* instead of *textual reviews*. Handling textual reviews is beyond our scope but the research scope of the Natural Language Processing (NLP) area.

The current approaches (Golbeck, 2005; Teacy et al., 2006; Zhang and Cohen, 2008; Yang et al., 2009) for opinion evaluation generally suffer from two main shortcomings:

1. Some of the previous approaches have been proposed to address the problems by using *social networks* (Golbeck, 2005) where opinions coming from *friends* in a user's social network are greatly valued. However, being friends of the user does not necessarily always share the same subjectivity (i.e. preference) with the user. In addition, computational trust models (Teacy et al., 2006; Zhang and Cohen,

2007; Yu and Singh, 2003b; Hang et al., 2009, 2013; Fang et al., 2012a,b) in the multi-agent system area evaluate the quality of opinions by designing agents for users to model the trustworthiness of advisors on the basis of their past behavior. The underlying assumption here is that a more trustworthy advisor could provide opinions of higher quality. However, they mainly focus on the dishonesty problem (e.g. unfair ratings) (Teacy et al., 2006) and merely on the user subjectivity difference (Fang et al., 2012b), or could not well distinguish dishonesty and subjectivity from each other (Noorian et al., 2011; Teacy et al., 2012). Besides, most of them strive to model an advisor's trustworthiness for a user on the basis of either the common experience between the user and the advisor (Teacy et al., 2006) or the whole history of experiences (Zhang and Cohen, 2007). Consequently, these models, on the basis of machine learning techniques, might fail since: (1) the user and advisor have limited past experience in a community, especially limited common experience towards same entities; and (2) a user's behavior is dynamic and evolving, which greatly increases the difficulty of tracking her behavior.

2. The social attributes that are related to the subjectivity and dishonesty of users have often been ignored in previous studies. Subjectivity and dishonesty, regarded as either the extrinsic or intrinsic characteristics of users (human-beings), are more related to the behavioral and psychological perspectives of users (Hosmer, 1995; Bond Jr et al., 2013) other than merely the statistical probability of data points prevalently adopted by Computer Science (Liu, 2010; Park et al., 2012). In other words, it is expected to understand users' motivation and needs of providing opinions when evaluating them: which kind of user subjectivity can lead to opinion difference between users, why some advisors would like to provide dishonest opinions, and how the subjectivity and dishonesty can be represented and distinguished by users' rating behavior? It is thus worthwhile to incorporate social factors (e.g. benevolence, integrity and competence) when addressing these two aspects for opinion evaluation.

With the above two points in mind, it is thus worthwhile to develop novel methods for opinion evaluation that consider the subjectivity and dishonesty issues, and simultaneously take the social factors into consideration.

1.2 Research Scope and Solutions

The primary goals of this dissertation are twofold: (1) to solve the subjectivity and dishonesty problems for opinion evaluation by designing trust models; and (2) to simultaneously take social factors into consideration when designing trust models. Particularly, to achieve these two goals, we conduct four studies which can be categorized into three parts: (1) social theory adaption for trust modeling; (2) social factors concerning subjectivity; (3) social factors concerning both subjectivity and dishonesty. Below, we elaborate each of these three parts.

1.2.1 Social Theory Adaption for Trust Modeling

As described in the previous section, trust models are regarded as effective methods for opinion evaluation in multiagent systems. However, the current approaches generally fail to capture the dynamic process through which a person's trustworthiness or her trust towards other users is cultivated in a system. Moreover, they also ignore the dependency relationships among users, or between users and the virtual environments. On the other hand, these basic issues are well formalized by the social dynamics theories ([Durlauf and Young, 2004](#)) in Social Science, one of which is the well-known diffusion theory ([Valente, 2005](#)). This motivates our first study to develop the DiffTrust ([Fang et al., 2013c](#)) model. In particular, we specifically employ Strange and Tuma's individual level-oriented heterogenous diffusion model ([Strang and Tuma, 1993](#)), considering that: (1) an advisor's trustworthiness in the view of a specific user may be different from the perception of other users (i.e. subjectivity); and (2) it is specifically sensitive to the interactions between the advisor and the user, and also subject to the user's own intrinsic nature.

Our DiffTrust model emphasizes the dynamic and evolutionary characteristics of trust, and highlights that the trustworthiness of an advisor may be perceived differently by different users, dependent on the environment, and embedded with a specific context. In DiffTrust, trust is recognized as a diffusive concept ([Williamson, 1993](#)), and an advisor's trustworthiness is regarded as an innovation in the diffusion theory. 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.

1.2.2 Social Factors Concerning Subjectivity

It cannot be denied that the subjectivity characteristic is more related to behavioral and psychological perspective of a person, i.e. as the social attribute. In this case, it is worthwhile for researchers in Computer Science to carefully examine how the subjectivity of users' perspectives can take effect in the process of providing ratings. Only in this way can meaningful and effective algorithms or models be developed. With this motivation, we carefully analyze the scenario of a user rating an entity from both psychological and behavioral perspectives, and identify two kinds of subjectivity that play a role in this process.

- Intra-attribute subjectivity: the subjectivity in evaluating the same attribute. When the user evaluates her satisfaction level with the entity, she considers each attribute related to that entity. Although the information about each attribute is objective, the evaluation (satisfactory level) of the attribute may be subjective.
- Extra-attribute subjectivity: the subjectivity in evaluating different attributes. When the user evaluates her satisfaction level with the entity, she may consider some attributes related to the entity more heavily than the rest.

These two types of subjectivity together contribute to the subjectivity difference among users when providing ratings to same entities. We thus propose the SARC (subjectivity alignment for reputation computation) (Fang et al., 2012b) model to address them. In particular, we employ the *Bayesian learning approach* to model each user's intra-attribute subjectivity, and *regression analysis* to learn the extra-attribute subjectivity on the basis of each user's past experience. Ratings provided by one user can thus be aligned for another user according to the two users' subjectivity. In this way, the subjectivity problem gets resolved by the SARC model.

1.2.3 Social Factors Concerning Subjectivity and Dishonesty

Although both the subjectivity and dishonesty are related to social attribute of users, they differ from each other in the sense that, the former one is an extrinsic (unconscious) characteristic. Only the difference in perspectives (i.e. subjectivity) between a user and an advisor makes the advisor *unintentionally* provide different ratings from the user. In contrast, dishonesty is an intrinsic (conscious) characteristic of the advisor, meaning that if the advisor is dishonest, she will provide an untruthful rating to the entity regardless of the user's characteristic. Therefore, it is necessary to clearly distinguish these two aspects.

We propose two models, i.e. PGTM (Fang et al., 2013a) and SubGroup (Fang et al., 2014), to clearly distinguish between dishonesty and subjectivity difference. PGTM is a novel probabilistic graphical trust model that computes the advisor's trustworthiness based on ratings and available trust relationships (if any). In PGTM, we adapt the well-known trust framework in Social Science (Mayer et al., 1995; McKnight and Chervany, 2001), and model the factors of advisors' intrinsic nature (dishonesty, i.e. benevolence, integrity, and competence), users' propensity to trust advisors, and difference in perspectives between users and advisors, as latent variables that may influence users' trust towards advisors.

PGTM distinguishes subjectivity and dishonesty by modeling them using different

sources of rating information, and captures their relationships with trust through the influence of chains in the probabilistic graphic model. Thus, when rating information is sparse, this model may not be very effective. Besides, it ignores the fact that dishonesty and subjectivity are overlapped with each other to certain extent. Alternatively, our SubGroup model addresses this problem by employing clustering techniques to model each advisor as part of groups. Particularly, we come up with several propositions in the literature or by intuition, which indicate features for the clustering analysis to distinguish subjectivity and dishonesty in users' rating behavior. The values of these features are expected to mirror the change of users' behavior over time, and thus the clustering analysis can also well tackle users' dynamic and evolving behavior. The clustering analysis consists of two layers. In the first layer, on the basis of indicative features extracted from users' rating behavior for distinguishing the subjectivity and dishonesty aspects, we employ DENCLU ([Zhang et al., 2005](#)) that crisply clusters advisors into different subjectivity groups and dishonest types. In the second layer, we adapt a fuzzy process to softly smooth and justify the clustering results of the first layer. Each advisor is assigned to two groups with respective membership degrees. Moreover, given the clustering results of the advisors (either subjective or dishonest), we further propose a simple alignment approach to help each user align ratings contributed by advisors to those of her own.

1.3 Research Contributions

The main contributions of this dissertation are summarized as follows.

1. Our DiffTrust model, stemmed from the well-known diffusion theory in Social Science, mainly deals with advisors' dishonesty. It is a novel trust model to evaluate the opinions of advisors by modeling their trustworthiness. It links the concepts in trust modeling with those in diffusion theory. In this way, DiffTrust can capture the dynamics of trust, and its dependency on other users and the environment.

Experimental results demonstrate that DiffTrust can consistently perform better in both loosely-connected and well-connected environments. Further, it can also help model the trustworthiness of advisors for users who are new to the system.

2. The SARC model addresses two kinds of subjectivity between users: intra-attribute subjectivity and extra-attribute subjectivity. It is a subjectivity alignment approach for reputation computation in reputation mechanisms when aggregating numerical ratings (opinions) provided by users towards the same entities. Although SARC only considers subjectivity, it is little affected by dishonest users, as validated in our experiments. Experimental results in a simulated e-commerce environment also verify that (1) SARC can more accurately and stably model sellers' reputation; (2) it is capable of coping with environments with dynamic buyer and seller behavior; (3) its requirement for successful implementation is not very strict.
3. Our PGTM and SubGroup models can distinguish and cope with dishonesty and subjectivity for opinion evaluation. In PGTM, we successfully adapt a well-known trust framework from Social Science, and model different social factors as latent variables that may influence users' trust towards advisors. Experimental results indicate that the latent variables in our model are both theoretically reasonable and computationally effective, and dishonesty and subjectivity are successfully distinguished.

The SubGroup model addresses a limitation of PGTM: the fact that dishonesty and subjectivity overlap with each other to certain extent. It is a two-layered clustering approach that models each advisor as part of groups, which can address the data sparsity problem. We conduct experiments on both a simulated distributed environment and real data (considered as centralized environments), and results verify that our approach can achieve better performance in comparison with the state-of-the-art trust models and recommender algorithms. These experiments also demonstrate that users still could extract valuable information from the opinions

contributed by some dishonest advisors.

4. As a new attempt to consider social factors in the trust area, our approaches contribute to bridging the gap between computational trust in Computer Science and psychological and behavioral trust in Social Science. Besides, it will induce more attention towards this direction. Moreover, we also examine the possibility of adopting these social factors (e.g. benevolence, integrity and competence) into other areas, such as recommender systems.

1.4 Dissertation Outline

The rest of this dissertation is organized as follows.

In Chapter 2, we review the related literature. Specifically, we begin with the introduction of online opinions and its social influence on users' decision making, as well as the involved quality problems. Then, we review the existing methods for opinion evaluation, including reputation mechanisms, computational trust models in multiagent systems, and recommender systems.

In Chapter 3, we present our DiffTrust model by beginning with an overview of the study, and then describe in detail the rationale of deriving the DiffTrust model from the diffusion theory. Next, we provide the computational procedures for users to model the trustworthiness of advisors for evaluating their opinions in DiffTrust. Finally, we show the experimental evaluation of our model on four real datasets, to verify the effectiveness of our model on trust modeling and opinion evaluation in comparison with four state-of-the-art and competing approaches.

In Chapter 4, we present our SARC model by beginning with an overview of the SARC model, including notations and basic procedures. Next, we describe in detail how our model learns users' intra-attribute and extra-attribute subjectivity, and aligns ratings provided by advisors. Finally, we show the experimental evaluation of our model in a simulated e-commerce environment where we compare our model with state-of-the-art

approaches in various scenarios, including basic, deceptive and dynamic environments.

In Chapter 5, we present our PGTM model by beginning with an overview of the study, and then describe the details of our model, including parameters, the generative process, model inference and parameter estimation, and trust prediction. Finally, we report on the experimental evaluation of our model on three real datasets by comparing with two competing approaches.

In Chapter 6, we present our SubGroup model by beginning with an overview of the study, and formally examine the subjectivity and dishonesty problems given users' rating behavior. Then, we describe the procedural framework of our model, and give the details of each procedure, including feature identification, cluster analysis and group alignment. Finally, we show the experimental evaluation of our model for both the task of opinion evaluation by comparing with competing trust models in a simulated distributed environment, and the task of recommendation by comparing with representative recommender algorithms on three real datasets.

In Chapter 7, we compare our four models and briefly explore the potential of our models in coping with other challenging scenarios such as resisting attackers. We then discuss how the social factors can be applied in other scenarios, such as recommender systems.

In Chapter 8, we conclude the dissertation by summarizing the studies and recapping their contributions, and also providing possible directions for future work.

CHAPTER 2

Literature Review

In this chapter, we provide a review on previous studies related to the research in the dissertation. We first introduce online opinions and their social influence in various areas, as well as the opinion quality problems in online communities. We then review the major existing methods related to our studies, including reputation mechanisms, computational trust models in multiagent systems, and recommender systems.

2.1 Online Opinions

With the advent of Web 2.0, users have been particularly productive in providing opinions (online word-of-mouth, e.g. online ratings and textual reviews) of virtually every entity ranging from televisions, cars, hotels, restaurants, books and movies to mobile apps, opinions and people ([Dellarocas, 2005](#)). Previous research on online opinions mostly concentrates on six aspects: connotation ([Hennig-Thurau et al., 2004](#)), motivation ([Picazo-Vela et al., 2010](#)), content ([Ye et al., 2007](#); [Korfiatis et al., 2012](#)), effect ([Chevalier and Mayzlin, 2006](#); [Chen et al., 2008](#); [Utz et al., 2012](#)), management ([Hennig-Thurau et al., 2004](#); [Park and Kim, 2008](#)), and individual differences ([Resnick and Zeckhauser, 2002](#); [Doh and Hwang, 2005](#); [Hu et al., 2011b](#)). These aspects almost cover the whole life cycle of online opinions. Below, we mainly focus on two particular issues: the influence of online opinions on users' decision making, and problems of online opinion quality.

2.1.1 Social Influence on Decision-making

Our society is increasingly relying on the digitized, aggregated opinions of others to make decisions (Muchnik et al., 2013) in various online communities, such as e-commerce sites, social networks and political forums. The power of word-of-mouth communication (i.e. opinions) and its influence on user decision making are well established in academic literature (Steffes and Burgee, 2009).

In e-commerce sites, the online opinions contributed by users towards products, compared with those provided by merchandisers themselves, convey more consumer-oriented product information about purchase and usage experience (Bickart and Schindler, 2001). On the other hand, it has long been recognized that these opinions are of great use to prospective users of these products (Lee et al., 2008), leading to substantial efforts being devoted to harnessing them. Resnick and Zeckhauser (2002) point out that consumers' online opinions have a direct effect on transactions. Chevalier and Mayzlin (2006) demonstrate that online opinions have a significant influence on book sales on Amazon US (the largest e-commerce website in the United States), while Fang et al. (2013b) find that on Dangdang (dangdang.com, the largest online book store in China), some aspects (e.g. the number and the ratings) of online opinions are also positively correlated with product sales. Online merchandisers have begun to embed consumers' online opinions in product advertisements as complements to product information (Lee et al., 2011).

Due to the significance of online opinions, online rating systems and online review systems, where all the activities (i.e. opinion contribution and opinion harnessing) happen, play a crucial role, especially on product sales. For example, Zhang et al. (2010) address the question of whether consumer-written opinions and professional editor-written opinions have different influences on the online consumers' purchasing intention. Based on real data collected from dianping.com (one of the leading consumer advice websites in China), they find that, in the Chinese context, consumers' ratings on quality of food, environments and services of restaurants and the number of online opinions

are positively correlated with the online popularity of restaurants, whereas professional reviews have a negative relationship with consumers' purchasing intention. Meanwhile, researchers use many methods to explore the correlation between product sales and different aspects of online opinions. [Chevalier and Mayzlin \(2006\)](#) declare that online opinions have a positive effect on book sales of both Amazon and Barnesandnoble.com. High quality of online opinions can improve sales, and opinions of rating 1 (i.e. low rating) have a more significant effect on product sales than opinions of rating 5 (i.e. high rating). Based on Transaction Cost Theory and Uncertainty Reduction Theory, [Hu et al. \(2008\)](#) use a combinative analysis to explore the effect of online opinions on sales of Amazon videos and books. [Ghose and Ipeirotis \(2011\)](#) use an econometric approach, text mining and the predictive control model to analyze the correlation between product sales and characteristics of online opinions including average helpfulness of historical opinions, information richness and readability. Besides, [Dellarocas et al. \(2007\)](#) try to figure out how to predict movie box office sales based on historical opinions.

Other than the social influence on consumers' decision making, online opinions are also increasing the impact on other perspectives of people's daily life. For example, [Steffes and Burgee \(2009\)](#) find that university students seeking information on which professor to take, weight the information they obtained from the online opinions contributed by others in online communities to be equally influential in their decision as their own primary experience with a professor. Online communities are considered as democratizing medium where people can access, create and exchange their opinions, and thus bring more citizens into the political process, especially young people ([Delli Carpini, 2000](#)). [Tumasjan et al. \(2010\)](#) further reveal that the online opinions created by users in Twitter plausibly reflect the offline political landscape, and thus can predict elections.

On the other hand, the pervasiveness of opinions in online communities induces the growing number of *opinion leaders* (i.e. users with a high level of opinion-giving behavior), and further increases their influence on other users' attitudes and behavior. For example, [Van Eck et al. \(2011\)](#) validate that online opinion leaders increase the speed of

the information stream and the innovation adoption process.

2.1.2 Quality Problems

Online opinions are expected to reflect advisors' experience with entities without bias, whereas it is not true in the real world. It is uncertain that in online communities, the extent to which online opinions are truthful "user-generated" content or merely comments produced by interested parties to promote their entities or demote those of competitors (Hu et al., 2011a,b, 2012; Xu et al., 2013). In online stores, some reviews are found to be provided by users without even using the products (Jindal and Liu, 2008) or consuming the services (Feng et al., 2012b). Besides, the problem of unfair ratings (unfairly positive or negative ratings) is fundamental and exists in almost every reputation system (Zhang and Cohen, 2006).

The opinion quality is not a hypothetical phenomenon. On eBay, a well-known example is that in 1999, three men highly rated each other and later sold a fake painting for a very high price (Jøsang and Ismail, 2002). In December 2002, AuctionBytes conducted a survey on the unfair rating problem on eBay and found that 39% of respondents felt that rating retaliation was a very big problem on eBay, 19% of respondents had received retaliatory ratings within the 6 months prior to the survey, and 16% had been a victim of rating extortion within that period³. On Taobao, some users figure out that stores selling fake products receive high reputation of 4.8/5, which is 20% higher than the average reputation of other stores selling the same kind of but real products⁴. When Amazon.ca (Amazon's Canadian website), due to software errors, unexpectedly revealed some true identities of its book reviewers, people found that a considerable proportion of online reviews are written by the books' own publishers, authors and their friends or relatives (Gurun and Butler, 2012). Hu et al. (2011a) found that fraudulent behavior in providing opinions is extremely serious for four kinds of scenarios in Amazon: non-

³<http://www.auctionbytes.com/cab/abn/y03/m09/i17/s01>.

⁴<http://bbs.taobao.com/catalog/thread/154504-256763011.htm>.

bestseller books, books for which opinions given are classified as not very helpful, books that experience greater variability in the helpfulness of their online opinions, and popular books as well as high-priced books. In the music industry, professional marketers are hired to surf various chat rooms and fan sites to post positive comments about new albums (Majumdar et al., 2007). Fake opinions also appear in the hospitality industry centered around hotels and restaurants (Hu et al., 2012). TripAdvisor faced legal action from up to 700 hotels which complained about being victims of ‘unfair criticism’ (i.e. unfair ratings) on TripAdvisor in 2010⁵. In the film industry, the problem of unfair ratings about movies is also a big issue, where on IMDB (www.imdb.com), some complainers listed 52 movies unfairly rated lower than 6/10 stars in 2012⁶.

It is thus essential to develop effective mechanisms to appropriately address this problem, and fairly evaluate the quality of opinions in order to determine how much to rely on them. As indicated before, this dissertation mainly focuses on ratings. Therefore, in the following section, we will review existing work related to our studies, also dealing with ratings instead of online reviews⁷.

2.2 Studies Related to Opinion Evaluation

In this section, we review three types of existing studies related to our work in the dissertation. Specifically, the reputation mechanisms are related to our SARC model in Chapter 4, the trust frameworks and trust propagation models are related to our diffTrust model in Chapter 3, the trust models for opinion evaluation are related to all the four studies, and the recommender systems are related to the SubGroup model in Chapter 6.

⁵<http://reputationonline.co.uk/2010/09/29/tripadvisor-to-face-legal-action-over-unfair-reviews/>.

⁶<http://www.imdb.com/news/ni23307748/>.

⁷There are a lot of previous studies (Xu et al., 2013; Jindal and Liu, 2008) in the data mining area working on detecting spam reviews using machine learning and data mining techniques. Since they are beyond our scope, we will not mention them in this dissertation.

2.2.1 Reputation Mechanisms

Reputation mechanisms allow users to model trustworthiness (i.e. reputation) of entities. In reputation mechanisms ([Resnick and Zeckhauser, 2002](#)), users who previously interact with an entity share their experience, normally in the form of a numerical rating reflecting the level of satisfaction for the experience with the entity. These ratings are aggregated to represent the entity's reputation. Other users can rely on the reputation value to make decisions on which entities to interact with. Reputation systems are particularly useful for users who have no or very little experience with entities.

In recent years, a great deal of research have been carried out on reputation mechanisms ([Resnick et al., 2000](#); [Jøsang and Ismail, 2002](#); [Teacy et al., 2006](#); [Resnick and Zeckhauser, 2002](#); [Swamynathan et al., 2010](#); [Coetzee et al., 2014](#)) in online communities, and have achieved a huge success, while one of those well-known reputation systems is run by eBay (www.ebay.com). eBay's reputation system, also as one of the earliest online reputation systems, gathers opinions from buyers of each transaction in the simple form of numerical ratings together with a short text description. It had some obvious drawbacks in its infancy such as opinions always being positive (and hence less distinguishable), reciprocal buyer and seller ratings, and hence amenable to trust prediction ([Resnick and Zeckhauser, 2002](#)). However, as it has matured, the eBay market rewards sellers with higher reputation values who have accumulated a lot of positive opinions. The reputation system now exhibits good robustness and sellers' reputations are positively correlated with product prices ([Resnick et al., 2006](#)). There are other successful commercial and live reputation systems ([Jøsang et al., 2007](#)), such as expert sites like Askme and Advocate, products review sites like Epinions and Amazon, Discussion Fora like Slashdot, Google's web page ranking system, supplier reputation systems and scientometrics- related sites.

At present, some online reputation systems (i.e. online opinion systems) have added the function that allows users to rate the opinions contributed by other users. For example, in Amazon, users can vote for others' opinions being either helpful or not. However, due to the unattractiveness of the design and the lack of effective incentive mechanisms, users

are reluctant to engage in the voting mechanism ([Fang et al., 2013b](#)).

2.2.2 Computational Trust Models in MAS

In this section, we review three parts of computational trust models in MAS: trust frameworks, trust propagation models, and trust models for evaluating advisors' trustworthiness considering the subjectivity and dishonesty issues.

Trust Frameworks

Numerous trust frameworks have been proposed to address trust issues in online communities (e.g. multi-agent systems). They can be mainly categorized into three types according to the incorporated information sources for modeling trust: (1) direct interactions, to model trust based on a user's direct experience with another user ([Marsh, 1994](#)); (2) indirect interactions, to model a user's trust based on other users' experience with this user; and (3) both direct and indirect interactions. Most of the frameworks can be classed into this third category ([Zacharia et al., 2000](#); [Sabater and Sierra, 2002, 2001](#); [Huynh et al., 2006](#); [Falcone and Castelfranchi, 2001](#); [Pinyol et al., 2012](#)), and we further explore the details of this category in view of its similarity with our DiffTrust framework.

For example, [Zacharia et al. \(2000\)](#) propose trust frameworks *Sporas* for loosely connected online communities and *Histos* for well-connected communities respectively. In these two models, a user's trust will never fall below an initial value, and users with higher trust will get smaller update. They also address the time decay effect of ratings by considering only the most recent ratings. However, *Sporas* and *Histos* may not be so general since they emphasize the trust between buyers and sellers. Besides, they fail to show the dynamic process of trust building since they only consider the time decay effect of ratings, but ignore the time decay effect of trust values.

Regret ([Sabater and Sierra, 2002](#)) and FIRE ([Huynh et al., 2006](#)) are two relatively complete trust frameworks, and incorporate similar elements. Regret takes into account information sources of three dimensions: individual dimension, social dimension and

ontological dimension. FIRE considers four kinds of information sources: direct interactions, role-based relationships, witness information and certified reputation. However, they still have following common limitations: (1) Regret assumes the existence of social relationships among users while FIRE recognizes the known role-relationships, but these relationships are not available in most real world cases (e.g. eBay); (2) the combination of different kinds of information sources are quite intuitive, but lack theoretical support; and (3) they are not adaptable to highly dynamic environments.

Several cognitive trust frameworks ([Falcone and Castelfranchi, 2001](#); [Pinyol et al., 2012](#)), which treat trust as subjective properties, have also been proposed. For example, the framework of [Falcone and Castelfranchi \(2001\)](#) indicates that trust is a mental state including three concepts: trust disposition, decision to trust and trusting behavior. The cognitive components that cause a user to trust another user are a set of beliefs, e.g., competence belief, disposition belief and dependence belief. The framework is hardly implementable. Based on this framework, [Pinyol et al. \(2012\)](#) integrate cognitive model (Repage) into a cognitive BDI agent. They define a belief logic to capture the Repage information, which can be used to build a complete graded BDI agents as a multi-context system to perform BDI reasoning. This approach can be called a computational model, but it consists of many beliefs and is complex, and then not so effective in large systems. Besides, it only considers a part of the information to model the image and reputation in the Repage model.

Our DiffTrust framework advances current state-of-the-art trust frameworks from the following three perspectives: (1) DiffTrust is a theoretically-supported computational framework; it reasonably integrates different information sources; (2) DiffTrust is an evolutionary framework and can address the dynamic nature of trust. Therefore, It address the initial users (cold start) problem, and also is adaptive to highly dynamic environments; and (3) DiffTrust combines the temporal information and spatial information (both physical and social spatial information) into the framework.

Trust Propagation Models

The approach of [Yu and Singh \(2003b\)](#) adopts the concept of *referral networks* where users cooperate with each other to find the trustworthiness of advisors by searching a social network. The social network is built upon two dimensions: expertise and sociability. However, this approach suffers from the “rumor problem”: that supposedly different information about an advisor’s trustworthiness received by a user may, in fact, have come from the same source. It also fails to work when new users do not have established trust relationships with others, and so it is not appropriate for communities where users are loosely connected. [Jiang et al. \(2013\)](#) propose a novel multiagent evolutionary trust model (MET) where each user acquires trust network information from her advisor and generates a candidate trust network using evolutionary operators. In MET, only trust networks providing more accurate entity reputation shall survive to the next generation. However, MET cannot deal well with the scenario where users have a relatively large number of advisors. The SALE POMDP approach of [Irissappane et al. \(2014b\)](#) allows each user to query her advisors about the trustworthiness of other advisors. Based on the information obtained from querying, it further reasons about the trustworthiness of advisors ([Oliehoek et al., 2012](#)). Both MET and SALE POMDP also suffer from the “rumor problem”. On the contrary, in our DiffTrust model, a user only accepts other users’ direct evaluation on an advisor to avoid the rumor problem. It also has no strict requirement on the minimum amount of experience for new users because various types of information (e.g. advisors’ profile information) can also be utilized in our model.

Some other trust propagation methods ([Sabater and Sierra, 2002](#); [Hang et al., 2009, 2013](#)) have also been proposed. [Sabater and Sierra \(2002\)](#) present a one-level propagation method where a user can make use of other users’ direct evaluation on an advisor. This method also considers the user’s social relations with other users, such as competition, co-operation and trade, which are typically difficult to obtain in real-world environments. On the basis of a social network built from users’ direct evaluation on each other, [Hang et al. \(2009\)](#) design a new algebraic approach called CertProp to propagate trust and address

the rumor problem. They formally define three operators: aggregation, concatenation and selection for trust propagation. However, CertProp suffers from the *unreachable witness information* problem where a user may not know anything about another user (called the witness) who holds information about an advisor. This problem has further been addressed in their recent work of the *Shin* approach ([Hang et al., 2013](#)) by considering the difference on trust evaluation towards same advisors between the user and the witness. However, both CertProp and Shin, grounded on the trust transitivity theory, are generally criticized by their accuracy, since trust evaluated through long paths has a high probability of being inaccurate. To analyze further, those trust propagation methods only consider the user's own experience with the advisor if such an experience exists, which may lead to inaccurate trust evaluation when the user has only limited experience with the advisor or the advisor dynamically changes her behavior.

Conversely, our DiffTrust model flexibly adjusts the weight of the user's own experience and other users' evaluation on the advisor. The quality of the other users' evaluation on the advisor is judged through the concept of social proximity, which has been proven to be effective in diffusion theory ([Angst et al., 2010](#)). In addition, social proximity can also be modeled based on other types of viable information provided in different environments other than the buyer's own experience with the other users' evaluation.

Trust Models for Evaluating Advisor Trustworthiness

Different trust models have also been proposed in multiagent systems to model the trustworthiness of advisors for evaluating their opinions. Some of them, such as ([Michael et al., 2010](#); [Yu and Singh, 2003a](#); [Whitby et al., 2004](#); [Teacy et al., 2006](#); [Zhang and Cohen, 2007](#); [Liu et al., 2011](#); [Irissappane et al., 2014a](#); [Liu et al., 2014](#)), focus on addressing opinions of low quality intentionally provided by dishonest advisors. For example, some of the approaches filter out the ratings of some users (advisors) whose past ratings differ significantly from the ratings of all advisors ([Whitby et al., 2004](#)), the ratings of a particular user ([Teacy et al., 2006](#)), or the ratings of both ([Zhang and Cohen,](#)

2008; Irissappane et al., 2014a; Liu et al., 2014). To be more specific, Whitby et al. (2004) design a statistical filtering technique that excludes the ratings of an entity that are not in the majority of the ratings for the same seller. The TRAVOS approach (Teacy et al., 2006) addresses the problem of unfair ratings by accomplishing two tasks. The first task is to estimate the probability that an advisor's ratings are accurate based on the advisor's previous ratings. The second task is to discount the ratings according to their accuracy. The personalized approach of Zhang and Cohen (2008) measures the reliability of an advisor (as well as her ratings) from two aspects: (1) private reputation calculated by comparing the advisor's ratings with the user's personal ratings regarding the commonly rated entities; and (2) public reputation estimated by comparing the advisor's ratings and other advisors' ratings regarding all entities. Ratings from less reliable advisors will be filtered or discounted. The above mentioned filtering approaches mainly deal with binary cases (i.e. ratings with binary values), but cannot effectively resolve the unfair ratings with multi-nominal degrees. Towards this problem, the iCLUB approach (Liu et al., 2011, 2014) is designed to filter unfair ratings with multi-nominal scales. It adopts clustering techniques and considers both users' local and global knowledge about entities. Irissappane et al. (2014a) further propose a biclustering-based approach to filter out dishonest advisors by dealing with the multi-criteria scenarios where entities are evaluated on multiple criteria. These aforementioned filtering approaches generally suffer from the risk of losing or discounting some important information.

Our SARC approach does not filter or discount ratings provided by an advisor with different subjectivity. Instead, our approach aligns/converts the ratings of the advisor to those that can be directly used by users according to the users and advisor's subjectivity learned by their agents. Due to the ignorance of subjectivity difference between advisors and users, the approaches mentioned above may misuse some important information caused by subjectivity difference, rather than dishonesty.

Some other approaches (e.g. Koster et al. (2010), also including our SARC (Fang et al., 2012b)), although rare, only consider subjectivity difference between users and

advisors, but ignore the dishonesty of advisors. [Koster et al. \(2010\)](#) propose a trust alignment approach based on Channel Theory. In this approach, each agent computes its own user's trust evaluation patterns, a set of "Specific Rules for Alignments" (SRAs) consisting of a trust value and a set of attributes, based on the interactions towards the same sellers (i.e. shared interactions). Then, clustering and Inductive Logic Programming (ILP) are used to generalize the SRAs to obtain the "General Rules for Alignments" (GRAs). GRAs are used to align trust advice provided by advisors. Compared to our SARC model ([Fang et al., 2012b](#)), one shortcoming of these subjectivity alignment approaches is that they ignore the intra-attribute subjectivity difference between a user and an advisor. Another shortcoming is that they require the user and the advisor to have interacted with a set of same entities (shared interactions), which may not be the case in an online community environment with a large population of entities. In addition, these approaches generally offer limited flexibility for users to deal with the dynamic behavior of entities and dynamic subjectivity of advisors. In contrast, our SARC model aligns each rating provided by an advisor towards an interaction with an entity other than an aggregated trust value of the entity. In this way, our approach is not affected by entities' changing behavior. Our SARC model updates the learned subjectivity of advisors once a while to cope with the possible dynamic subjectivity of advisors. Our SARC model does not rely on shared interactions. Instead, the agent of each user makes use of the ratings and detailed reviews provided by the user about her interactions with any entities, to learn the user's intra-attribute and extra-attribute subjectivity. However, our SARC model, together with other trust alignment approaches, may mistakenly treat dishonest advisors as those having subjectivity difference with users.

An approach, that also requires users to provide detailed reviews of their interactions with entities to address the subjectivity difference problem, is POYRAZ ([Sensoy et al., 2009](#)). The POYRAZ approach models the reputation of entities on the basis of detailed reviews containing values of the objective attributes of interactions with entities, rather than numerical ratings. However, this approach requires users (advisors) to always

provide a detailed review for each interaction, which is time-consuming. In contrast, our SARC model requires users to provide detailed reviews at the beginning of interacting with the reputation system. Afterwards, detailed reviews are required only if there is a need to update the learned subjectivity of users (advisors).

A few approaches have also been proposed to model both dishonesty and subjectivity of advisors, which are related to our PGTM and SubGroup models. BLADE ([Regan et al., 2006](#)) 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. HABIT ([Teacy et al., 2012](#)) extends BLADE by additionally considering third party information, but it might suffer from the same problem as BLADE. Prob-Cog ([Noorian et al., 2011](#)) 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 being dishonest about the user. In consequence, advisors having large subjectivity difference with the user will be misclassified as dishonest. PRep ([Haghpanah and desJardins, 2012](#)) learns advisors' behavior using Bayesian learning and then adjusts their opinions (biased or unbiased) according to the learned types of advisors. However, in common with Prob-cog and BLADE, it would wrongly treat some dishonest users as subjective ones, or vice versa. Besides, almost all the aforementioned approaches have strict requirements on the number of available interactions between users, which will be affected by the data sparsity problem where users may not have interacted with many common entities.

In contrast, our PGTM 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. It should be noted that, PGTM 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. However, PGTM ignores the fact that dishonesty and subjectivity overlap with each other to a certain extent from the perspective of information sources.

Our SubGroup model distinguishes subjective users from dishonest ones. Further, it addresses the data sparsity problem by considering advisors as part of groups. In contrast with those discounting or filtering approaches, ratings of advisors in SubGroup are aligned to the user's own so as not to lose information. Moreover, the set of features used to represent advisors can capture their dynamic and evolving behavior.

2.2.3 Recommender Systems

Social network based methods, e.g. trust-aware recommender systems where users' information sources can be enriched by past experience or recommendations of trust neighbors (advisors), are designed with the assumption that an advisor in a user's social network could provide more reliable opinions to the user. For example, Massa and Avesani (2007) replace user similarity with explicitly specified trust relationships, and also allow trust relationships to propagate through the trust networks. They show that more robust recommendations can be produced without significant loss in accuracy. Golbeck (2006) introduces a trust-flow-based method (called *TidalTrust*) to compute rating predictions for target items. She finds out that better accuracy can be achieved. Later works (Chowdhury et al., 2009; Ray and Mahanti, 2010) claim that better performance can be obtained by integrating both trust and similarity for recommendations. Jamali and Ester (2009) design the *TrustWalker* approach to randomly select neighbors in the trust network formed by users and their trusted neighbors. TrustWalker combines trust information of the selected

neighbors with an item-based technique, where both the ratings of the target item and similar items are considered. The recent work conducted by [Guo et al. \(2012\)](#) focuses on the problems of data sparsity and cold start from which traditional recommender systems suffer. They empirically contend that by merging the ratings of trusted neighbors, the preferences of active users can be better modeled and hence the performance is improved. However, a generally agreed proposition states that friends may not share similar preferences ([Jøsang et al., 2011](#)). In other words, friends (trust neighbors) of a user might be intrinsically honest, but not necessary share the same subjectivity with the user.

In addition to these memory-based approaches, model-based approaches are also employed in recommender systems. Approaches related to our SubGroup model include clustering-based approaches, and matrix-factorization ones. Clustering-based methods reduce the search space in recommender systems by employing clustering techniques such as the weighted co-clustering ([George and Merugu, 2005](#)) method to cluster similar users or entities, and hence ratings of clustered users or entities are integrated to make predictions for corresponding users. However, they only employ information about users in the same cluster, while in our SubGroup model information about users from other clusters can also be effectively incorporated. Matrix factorization has become popular recently in recommender systems. It fits the user-entity rating matrix using low-rank approximations, i.e. user-feature matrix and entity-feature matrix, and employs low-rank matrices to provide further predication. For example, [Mnih and Salakhutdinov \(2007\)](#) propose a probabilistic matrix factorization model by assuming Guassian observations on observed user-entity ratings. [Ma et al. \(2008\)](#) design a latent factor model called *SoRec* based on probabilistic matrix factorization. They fuse the user-item rating matrix with a user-user trust matrix by sharing a common latent low dimensional user feature matrix. The two matrices are factorized by three sets of latent features: user vector and feature vector (for each user), and item vector. Experimental results demonstrate that *SoRec* outperforms the basic matrix factorization model and other trust related

neighborhood models. However, although trust information is considered, real world recommendation processes are not reflected, where the two sets of latent features for each user cause the low interpretability of the model. To overcome this problem and model trust-aware recommender systems more realistically, they further propose *RSTE* (Ma et al., 2009), a linear combination of a basic matrix factorization technique and a trust-based approach. Jamali and Ester (2010) later enhance this model by enabling trust propagation in their *SocialMF* model. However, all the above-mentioned model-based approaches (without using trust) ignore the dishonesty problem when considering other users' recommendations, while the ones considering trust information suffer from the same problem measured in the trust-aware recommender systems.

In recommender systems, the problem of shilling attacks has also increasingly caught researchers' attention (Lam and Riedl, 2004; Chirita et al., 2005; Gunes et al., 2012). Shilling attacks can be divided into two types according to their intent: push attack that aims to increase the popularity of targeted entities by providing lots of positive ratings, and nuke attack that intends to decrease the popularity of targeted entities by giving lots of negative ratings. Generally speaking, the memory-based approaches (without trust) are shown to be vulnerable to shilling attacks (O' Mahony et al., 2002a,b), while the model-based approaches (without trust) are demonstrated to have relatively better performance in terms of shilling attacks compared to the memory-based approaches. Trust-aware recommender systems are validated to be more robust than the others (Ji et al., 2007). Some approaches designed to resist shilling attacks have also been particularly proposed in recommender systems. For example, O' Mahony et al. (2003) develop a neighborhood filtering scheme where clustering is employed to place the ratings into two groups (i.e. malicious and genuine), and the prediction is estimated only using ratings from the genuine group. In this case, it also has the same problem as filtering approaches to computational trust mentioned earlier. Zhang and Xu (2007) propose a topic-level trust-based prediction algorithm to resolve the shilling attacks. They compute the entity-level trustworthiness of each user based on her neighbors, and then estimate the topic-level trust

by averaging the entity-level trust of the user on the entities belonging to the same topic. Finally, they incorporate the topic-level trust into a traditional recommender scheme. This kind of trust-aware algorithm would suffer from the same problems mentioned in computational trust models and trust-aware recommender systems.

2.3 Summary

Users have been particularly productive in providing opinions of virtually every entity ranging from products or services in e-commerce systems to comments or information in social media. On the other hand, they are also increasingly reliant on user-generated opinions in online communities to make a variety of decisions. For example, in e-commerce systems, user-generated content can provide more consumer-related product information, and thus greatly promote product sales. In social media, young people are more involved in seeking information or comments generated by others to make decisions, such as choosing professors whose courses they would like to take in universities and judging the politicians for their election decisions.

The increasing importance of online opinions also raises users' concerns regarding the quality of these opinions. It has been revealed that many interested parties would provide unfair ratings or untruthful reviews to promote their own entities, or demote the entities provided by their competitors. It thus becomes crucial to design effective mechanisms to address the opinion quality problem.

In Section 2.2, we reviewed existing methods focusing on the information source of users' rating behavior. In particular, previous trust frameworks suffer from problems of incomplete information and not being able to track users' dynamic behavior. These problems are moderately addressed by our DiffTrust work. Of the trust propagation models, referral networks have the "rumor" problem, and fail to work when new users enter the system. In contrast, our DiffTrust model employs direct trust evaluations to avoid the rumor problem, and can deal with the new users. Besides, other trust propagation

methods, grounded on the trust transitivity theory, are criticized by their accuracy, while this problem is also properly tackled by our diffTrust model.

For the trust models that evaluate advisors' trustworthiness by considering subjectivity difference, or dishonesty, or both, they may either misuse some important information caused by dishonesty, or vice versa. They may wrongly treat subjective users as dishonest ones, or dishonest users as subjective ones. Our PGTM and SubGroup models address these issues by explicitly distinguishing subjectivity difference and dishonesty in trust modeling or opinion evaluation using machine learning techniques. Moreover, all of our four models take social factors into consideration, which are ignored by most existing approaches.

The trust-aware recommender systems have been proposed to address the problems by using *social networks* where opinions coming from *friends* in a user's social network are greatly valued. However, being friends of or trusted by the user does not imply they share the same subjectivity (i.e. preference) with the user. Further, model-based approaches (without using trust) generally ignore the dishonesty problem when considering other users' recommendations.

CHAPTER 3

DiffTrust: A Trust Model Stemmed from Diffusion Theory

Trust has been recognized as a diffusive concept ([Williamson, 1993](#)). 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 (called DiffTrust) from this well-studied theory by considering an advisor's trustworthiness as an innovation. 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 capture the *dynamics* and *subjectivity* of trust, and its *dependency* on other users and the environment.

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, such as providing a rating to the 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 of the advisor is computed by the agents of other users in the social network. The agent of the user considers another user's trust evaluation of 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.

We conduct experiments on real data obtained from eBay (www.ebay.com), FilmTrust (trust.mindswap.org), Epinions (www.epinions.com) and Flixster (www.flixster.com), to verify the effectiveness of our model in comparison with state-of-the-art approaches including TRAVOS ([Teacy et al., 2006](#)), the personalized approach ([Zhang and Cohen, 2007](#)), CertProp ([Hang et al., 2009](#)) and Shin ([Hang et al., 2013](#)). The results demonstrate that our model can more accurately model the trustworthiness of advisors than these other approaches.

The rest of this chapter is organized as follows. We describe the details of our DiffTrust model in Section 3.1. After that, we conduct experiments to verify the effectiveness of our approach in Section 3.2. Finally, we summarize this work in Section 3.3.

3.1 The DiffTrust Model

In this section, we describe in detail the rationale of deriving the DiffTrust model from diffusion theory, and provide computational procedures for users to model trustworthiness of advisors for evaluating their opinions.

3.1.1 Trust and the Diffusion Theory

Our DiffTrust model is inspired by diffusion theory ([Strang and Tuma, 1993](#); [Strang and Soule, 1998](#)) where the home territory is an innovation (e.g. technology, idea or object) ([Valente, 2005](#)). Diffusion theory seeks to explain how, why and at what rate a

new innovation spreads through a community. Rogers (1995) indicates that “diffusion is the process where an *innovation* is communicated through *certain channels* over *time* among the members of a *social system*”.

According to Williamson (1993), it is more appropriate to consider trust as a diffusive concept. In trust modeling, it is critical and more meaningful to explore the context within which trust is embedded, and to explore the processes through which trust is cultivated. Further, a user’s trust perception towards another user is not static but evolves as the context changes. That is, we can consider a user’s trustworthiness perceived by other users as a systematic process, involving the purposes of evaluating and using trust of the user (why), the ways and channels of inducing trust towards the user (how), and the induced trust degree (at what rate) varying over time for other users or other contexts. We can clearly see the similarities between trust and innovation in diffusion theory. Both of them are *dynamic* and *evolutionary* (varying over time), *subject* to perceiving (adopting) users (i.e. influenced by the intrinsic nature of these users), and *dependent* on the environment. It is thus natural to derive a trust model based on diffusion theory.

In a system, an advisor’s trustworthiness in the view of a specific user, which may be different from the perception of other users, is sensitive to the interactions between the advisor and the user and also subject to the user’s own intrinsic nature. Considering this, it is more appropriate to use the individual-level framework of diffusion theory. We specifically employ Strang and Tuma’s individual-level oriented heterogeneous diffusion model (Strang and Tuma, 1993). This model emphasizes the heterogeneous characteristics of both spatial and temporal information. Similarly, the diffusion of an advisor’s trustworthiness among other users should also consider the spatial heterogeneity among the other users, because different users have different chances of perceiving the advisor’s trustworthiness and different strength of affecting others’ perception. We are also concerned with the temporal heterogeneity among all historical interactions where a more recent interaction will have a greater effect on trust evaluation. In Strang and Tuma’s model, there are four main factors influencing the innovation adoption process of

Table 3.1: Summary of Notations in DiffTrust

Notations	Description
u	A user.
a	An advisor.
$c \in \{c_1, c_2, \dots, c_m\}$	The set of contexts in the system, while m is the total number.
U_a^c	The set of users that have adopted a in their social networks under context c .
$T(u, a, t, c)$	u 's trust on a at time t under c .
$DT(u, a, t, c)$	u 's direct trust on a at time t under c .
$CI(u, U_a^c, t)$	The contagious influence of users in U_a^c .

a potential adopter: *the susceptibility of the potential adopter* (**individual tendency**, an adopter's intrinsic nature to adopt a new innovation), *the infectiousness (contagiousness) of early adopters* and *the proximity of the pairs consisting of one early adopter and the potential adopter* (**contagious influence**, the impact from the users already adopting the innovation), *the characteristics of the innovation* (**direct connections**, the potential adopter's perception about the characteristics of innovations through direct observation or communication), and *temporal variation* as a function of time since adoption. We will explain next how our DiffTrust model is derived based on these factors.

3.1.2 The Trust Model

In DiffTrust, each user is equipped with a software agent being responsible for modeling the trustworthiness of advisors for its user. Each user has a set of past interactions with some entities and provides a rating in the range of $[0, 1]$ for each interaction. We assume that a user u and an advisor a have both previously interacted with a set of the same entities. Based on these shared interactions, the agent computes user u 's direct trust on advisor a . The direct trust information of every user will be used to form a social network of all users. Let us assume that the set of users who have adopted advisor a in their

Figure 3.1: The DiffTrust Model for Evaluating the Trustworthiness of Advisors

social networks under the context c is denoted by the set U_a^c . Our objective is to model the trustworthiness of advisor a perceived by user u at time t . The trustworthiness of the advisor is modeled with respect to each context c ($c \in \{c_1, c_2, \dots, c_m\}$, where m is the number of contexts in the system), at time t as $T(u, a, t, c)$. The major notations are summarized in Table 3.1.

We adopt the individual-level oriented heterogenous diffusion model (Strang and Tuma, 1993) to derive the DiffTrust model based on its four factors. More specifically, as shown in Figure 3.1, the agent of user u models advisor a 's trustworthiness by incorporating the following three major parts:

- *Susceptibility of user u trusting advisor a (individual tendency)*. This refers to user u 's intrinsic nature without any direct or indirect experience with advisor a . We consider two factors: user u 's static initial trust ($T_0^u \in [0, 1]$), and the number of u 's neighbors in the social network at a specific time. A larger number of neighbors implies that user u has already experienced a lot in the system. Therefore, she is more likely to be open to other unknown advisors and confident in her own trust evaluation.
- *Effect of direct connections between user u and advisor a (direct connections)*, representing user u 's direct observations on advisor a . Two factors are involved: direct trust based on shared interactions between u and a , and *social proximity* of a with respect to u , S_a^u , which is treated as a consequence of u 's direct observation

on a 's spatial information.

- *Contagious influence of set U_a^c (contagious influence)*. We consider the infectiousness of each user in U_a^c , by sharing their evaluation on advisor a derived from their shared interactions. We make use of their trust evaluations by computing the *social proximity* of each specific user in U_a^c with respect to user u . A socially closer user to u indicates a higher probability that u will trust the user's trust evaluation on a .

We also consider the *time decay effect* of shared interactions (temporal variation). The effect of previous shared interactions decreases as time goes by, and thus trust values computed based on long past evidence should also be decreased accordingly.

3.1.3 Trust Computation

Before introducing the specific computational steps, we first clarify some important concepts related to our model.

Related Concepts

Spatial information is used to model the *social proximity* between users. Social proximity, believed to be able to engender trust (Angst et al., 2010), can be flexibly modeled according to the consideration of identified spatial information. Two kinds of spatial information are considered in our model, *physically spatial information* and *socially spatial information*. The former refers to users' physical location and identity. It can be obtained from users' profile information. For this kind of information, the method for calculating context similarity can be used to model the social proximity. The latter refers to users' social identity and status, such as users's position in the social network and their neighbors. Traditional proximity metrics such as the number of common neighbors (CN) (see Equation 3.1), fraction of common neighbors, or Jaccard coefficient and the Adamic-Adar score can be adopted to measure the social proximity of two users (Lerman et al., 2012). Note that we prefer to use *socially spatial information* when both of the

two kinds of spatial information are available, because social proximity derived from socially spatial information is proven to be more infusive than that from physically spatial information (Angst et al., 2010).

$$CN = \frac{1}{2}(|\Delta| + |\Delta'|) \quad (3.1)$$

where for two users u and v : $\Delta = \Gamma_{\text{out}}(u) \cap \Gamma_{\text{in}}(v)$ and $\Delta' = \Gamma_{\text{in}}(u) \cap \Gamma_{\text{out}}(v)$. Here, $\Gamma_{\text{out}}(u)$ represents the set of out-neighbors of node u , which in our research corresponds to the set of users being directly trusted by u . Similarly, $\Gamma_{\text{in}}(u)$ represents the set of users having direct trust on u .

Temporal information is bounded with other information, such as trust values and past interactions between users, whose influence would be discounted over time. We mainly consider the time when an interaction happens, when a user is added to another user's social network, and when a trust value is computed.

Context consists of both spatial information and temporal information, according to *Situation Dependency Theory* (Figge, 2004). In our work, context is usually mentioned together with an interaction or an advisor's trustworthiness, i.e. the context of a specific *interaction*, and advisor a 's *trustworthiness* perceived by user u under a specific context.

Computational Steps

We now elaborate the computational steps in detail.

Step 1: Model of direct trust based on shared interactions. As indicated in Section 3.1.2, user u models the direct trustworthiness of advisor a based on their shared interactions. This is a continuous process. Whenever a new shared interaction happens between u and a , u 's trust on a needs to be updated based on the new observation. We assume that based on the previous shared interactions of user u and advisor a , at time t_0 , u 's direct trust towards a under context c is denoted as $DT(u, a, t_0, c)$. At time t , u and a have a new shared interaction with a same entity, denoted as i_u under the context c_u and i_a under the context c_a respectively. Note that contexts c_u , c_a and c may not be the

same. The problem can be identified as to update u 's direct trust towards a at time t as $DT(u, a, t, c)$ by considering the effect of the new shared interaction i_u and i_a .

We first model the direct influence ($DI(i_u, i_a, t)$) of the new shared interaction on the trustworthiness of advisor a . Intuitively, in the shared interaction, if advisor a 's opinion is more similar to that of user u towards the same entity, a can be considered as more trustworthy. Thus, $DI(i_u, i_a, t)$ is formulated as the similarity between i_u and i_a :

$$DI(i_u, i_a, t) = (1 - |i_u - i_a|) \times S(c_u, c_a) \times S(c, c_u) \quad (3.2)$$

where $S(c, c_u)$ is the similarity between contexts c and c_u , and $S(c_u, c_a)$ is the similarity between contexts c_u and c_a . Calculation of context similarity will be specified later. In this way, we align the user and advisor's interactions with the same entity to context c .

We then update user u 's direct trust towards advisor a based on the new shared interaction at time t under context c , by combining the new shared interaction (i_u and i_a) with the previous trust values $DT(u, a, t_0, c)$ at t_0 , as follows:

$$DT(u, a, t, c) = \frac{DT(u, a, t_0, c)\lambda^{(t-t_0)} + DI(i_u, i_a, t)}{1 + \lambda^{(t-t_0)}} \quad (3.3)$$

where $0 < \lambda \leq 1$ is a time decay factor for user u to decrease the effect of old shared interactions between u and a .

Note that two special scenarios need to be considered: (1) if there is no new shared interaction after t_0 , we consider the time decay effect of previously computed trust value at t_0 such that $DT(u, a, t, c) = DT(u, a, t_0, c)\lambda^{(t-t_0)}$; and (2) if there is no shared interaction between u and a till time t , the direct connections between u and a can be modeled by the social proximity of u and a instead, which could be considered as user u 's direct observation of advisor a 's characteristics, that is $DT(u, a, t, c) = S_a^u$.

Step 2: Model of contagious influence. According to diffusion theory, the contagious influence of users in U_a^c , denoted as $CI(u, U_a^c, t)$, can be modeled as the weighted average of trust evaluation on advisor a from each user $x \in U_a^c$, denoted as $DT(x, a, t, c)$.

Each weight corresponds to user x 's *social proximity* with respect to user u denoted as S_x^u . The way of computing social proximity can vary depending on the information available in the environment. In our experiments, we use Equation 3.1 to compute social proximity. Higher social proximity means that two users are more similar/closer according to their spatial information, and user x 's evaluation on advisor a is pitched to resonate more in the mind of user u . The contagious influence of users U_a^c can then be formalized as follows:

$$\text{CI}(u, U_a^c, t) = \frac{\sum_{x \in U_a^c} (\text{DT}(x, a, t, c) \times S_x^u)}{\sum_{x \in U_a^c} S_x^u} \quad (3.4)$$

Step 3: Combine the three factors. We now evaluate the trustworthiness of advisor a with regard to user u at time t under context c , $T(u, a, t, c)$, by considering the three factors: direct connections, contagious influence and susceptibility (initial trust) of u , as follows:

$$T(u, a, t, c) = \omega_1 \text{DT}(u, a, t, c) + \omega_2 \text{CI}(u, U_a^c, t) + (1 - \omega_1 - \omega_2) T_0^u \quad (3.5)$$

where ω_2 is the weight of the contagious influence of the set of users in U_a^c and can be modeled as the average social proximity of users in U_a^c with respect to user u under the context c , and ω_1 is the weight of the direct trust derived from shared interactions. The weight ω_1 is influenced by user u 's confidence in the system. It relates to the number of neighbors user u has. The more neighbors u has, the more experienced and more confident u will be in relying on her own evaluations. We calculate ω_1 as follows:

$$\omega_1 = \begin{cases} \frac{N_u^t}{N_{min}} & \text{if } N_u^t \leq N_{min}; \\ 1 & \text{otherwise.} \end{cases} \quad (3.6)$$

where N_u^t is the number of u 's trust neighbors in the social network at time t , and N_{min} is the minimum number of neighbors needed for user u to be totally confident about her own evaluation. We adopt the method in (Zhang and Cohen, 2007) to compute the value

of N_{min} according to an acceptable level of error ε (for u) and a confidence measure γ , as follows:

$$N_{min} = -\frac{1}{2\varepsilon^2} \ln \frac{1-\gamma}{2} \quad (3.7)$$

Note that $\omega_1 = 1$ is only used in direct trust part, and in order to more accurately model the advisor's trustworthiness, the agent will always consider direct trust and contagious influence together. That is to say, in Equation 3.5, if $\omega_1 + \omega_2 > 1$, then we change the weights to be: $\omega_1 = \frac{\omega_1}{\omega_1 + \omega_2}$ and $\omega_2 = 1 - \omega_1$.

When there is no shared interaction between user u and advisor a , a will not be included into u 's social network. If a is a newcomer of the system, we consider only the social proximity between u and a and u 's initial trust (without contagious influence). An advisor will be added into user u 's social network only when shared interactions between the advisor and the user are identified. However, if an advisor's trustworthiness from the view of user u based on shared interactions is equivalent to the initial trust T_0^u , the advisor will be immediately excluded from u 's social network. Intuitively, it represents that user u 's previous shared interactions with the advisor have been forgotten and thus lost influence on trust evaluation of the advisor. Then, only new shared interactions between the advisor and the user will be taken into consideration in the trust computation.

Context Similarity

The context information in our work is represented by an ontology, and we employ the existing largest ontology - the LinkedData ontology (Bizer et al., 2009) to present the concepts involved for describing the contexts in a specific system. We assume two contexts c_i and $c_j \in \{c_1, c_2, \dots, c_m\}$ represented by a set of concepts (objects) in our universe, and relations between the concepts are generalization relationships (e.g. *isA* and *partOf*). Contexts $c_i = \langle c_i^1, c_i^2, \dots, c_i^n \rangle$ and $c_j = \langle c_j^1, c_j^2, \dots, c_j^n \rangle$, where n is the number of concepts describing a context. We adopt the definition of concept similarity in (Radinsky et al., 2012) that two concepts (e.g. apple and pear) are similar if they relate to a third concept (e.g. fruit). More formally, the distance between two

concepts is defined as the length of the shortest generalization path between the two concepts. For example, the distance between apple and pear is 1 ($apple \xrightarrow{isA} \mathbf{fruit}$, and $pear \xrightarrow{isA} \mathbf{fruit}$), while the distance between apple and rice is 2 ($apple \xrightarrow{isA} fruit \xrightarrow{isA} \mathbf{plant}$, and $rice \xrightarrow{isA} vegetable \xrightarrow{isA} \mathbf{plant}$). We then define the similarity of two contexts c_i and c_j as:

$$S(c_i, c_j) = \frac{1}{\sum_{y=1}^n \text{dist}(c_i^y, c_j^y) + 1} \quad (3.8)$$

where $\text{dist}(c_i^y, c_j^y)$ is the distance between concepts c_i^y and c_j^y defined previously using the shortest generalization path. Note that social proximity from physically spatial information can also be modeled in this way because physically spatial information can also be described by the ontology.

3.2 Experiments

We carry out two kinds of experiments to evaluate the performance of our DiffTrust model. The first kind is to verify whether the trustworthiness of advisors modeled by DiffTrust can be used to accurately model the trustworthiness of entities in reputation systems. These experiments are conducted on real data collected from eBay, modeling trustworthiness of sellers for buyers to predict the outcomes of their future transactions with the sellers. The second kind is to verify whether the modeled trustworthiness of advisors is the same as what have been explicitly indicated by users. We conduct these experiments using real datasets extracted from FilmTrust, Epinions and Flixster. We also compare our model with several competing models:

- **TRAVOS**, proposed by [Teacy et al. \(2006\)](#), relies on an advisor's historical ratings to filter unfair ratings.
- **The personalized approach**, proposed by [Zhang and Cohen \(2007\)](#), considers an advisor's reliability from two aspects: private reputation (similar to direct trust) and public reputation (by comparing the advisor's ratings with other advisors' ratings).

- **CertProp**, proposed by [Hang et al. \(2009\)](#), defines three operators (e.g. aggregation, concatenation, and selection) to propagate an advisor's trustworthiness through a user's social network).
- **Shin**, proposed by [Hang et al. \(2013\)](#), also considers the difference on trust evaluation towards same advisors between the user and the advisor to address the unreachable advisor problem in CertProp.
- **Simple Average**, also called **baseline** in our paper, which judges a seller's trustworthiness for a buyer by aggregating the ratings from advisors without considering their trustworthiness.

Of all the benchmarks, TRAVOS and the Personalized approach are representative in the set of filtering approaches, while CertProp and Shin are representative of the trust propagation models (See Chapter 2 for details).

3.2.1 Data Acquisition

For eBay, we first randomly select 26,922 sellers selling products in different categories and collect all their past transactions (including 969,213 positive ratings, 1914 neutral ratings and 3590 negative ratings). The data is crawled from April 10, 2000 to June 4, 2011. Then, we randomly select 23,630 buyers from all the buyers who have previously interacted with at least one of the selected sellers. We collect all past transactions of these buyers. Each transaction consists of a buyer ID, seller ID, rating provided by the buyer, and the time of the transaction. We then select 3,046 target sellers who have at least one unsuccessful (negative or neutral) past transaction to fully test the performance of different models. Correspondingly, the number of buyers in the buyer set is 5,531. We then predict the outcomes of past transactions that buyers have conducted with sellers using the leave-one-out strategy. More specifically, to predict the outcome of a transaction between a buyer b and a seller s at time t , we select ratings provided by other buyers (advisors) towards seller s before time t . The trustworthiness of advisors is modeled using

Table 3.2: Statistical Information about the Four Real Datasets

Datasets	eBay	FilmTrust	Flixster	Epinions
Statistics	186,957 buyers	1,508 users	1,000 users	500 users
	3,046 sellers	2,071 items	2,867 items	44,288 items
	214,115 transactions	35,497 ratings	7,905 ratings	66,287 ratings
Data types	buyers→sellers	users→items users→users	users→items users→users	users→items users→users
Vertices in social network	5,531 (buyers)	1,508 (users)	1,000 (users)	500 (users)
Edges in social network	1,144,675	1,632	3,337	23,848
Rating scale	-1, 0, 1	0.5, 1, 1.5, ..., 4	0.5, 1, 1.5, ..., 5	1, 2, 3, 4, 5
Trust between users	-	1	1	1
Ratio of having shared interactions of all users	7.5%	81.9%	1.5%	21.8%
Maximal number of shared interactions between users	21	105	444	121

different approaches based on all ratings provided before time t . The trustworthiness of seller s from the point of view of buyer b can then be evaluated by aggregating ratings from advisors weighted by their trustworthiness. In the end, the outcome of the unknown transaction can be predicted using the computed trustworthiness of seller s .

The FilmTrust dataset (provided by [Guo et al. \(2013\)](#)) consists of 1, 508 users, 2, 071 movies and 35, 497 movie ratings issued by the users. There are 1, 632 trust relationships explicitly identified by users, which are used as test data in our experiments. These trust relationship are directed. That is, a user a trusting another user b does not imply b also trusting a . On the basis of shared interactions (commonly rated items) between users, we model the trust value between two users using different models. The same evaluation method has been used on the Flixster and Epinions datasets. We randomly select 1, 000 users from the Flixster dataset⁸, and 500 users from the Epinions dataset⁹. The statistical information of these four datasets is summarized in Table 3.2.

⁸<http://www.cs.sfu.ca/~sja25/personal/datasets/>.

⁹http://www.trustlet.org/datasets/downloaded_epinions.

3.2.2 Evaluation Metrics

We use three metrics to measure the performance of different approaches: the Matthews Correlation Coefficient (MCC) for the eBay dataset, precision for FilmTrust, Epinions and Flixster dataset, and the mean absolute error (MAE) for all four datasets. MCC is a measure for the quality of binary classifications. It is generally regarded as a balanced measure which can be used even if the classes have distinct sizes. Thus, it is very suitable for our eBay dataset, where over 95% of the historical transactions are positive. MCC can be calculated as follows:

$$MCC = \frac{t_p t_n - f_p f_n}{\sqrt{(t_p + f_p)(t_p + f_n)(t_n + f_p)(t_n + f_n)}} \quad (3.9)$$

where t_p , t_n , f_p and f_n refer to an actual positive transaction predicted to be successful, an actual negative transaction predicted to be unsuccessful, an actual negative transaction predicted to be successful, and an actual positive transaction predicted to be unsuccessful, respectively. A MCC value of 1 represents perfect prediction, 0 an average random prediction and -1 the worst possible prediction. For the trust value prediction in FilmTrust, Epinions and Flixster, there is only information about the relationship that a user trusts an advisor. Precision is thus used as the evaluation metric, which refers to the actual fraction of successful predictions of those trust relationships. More specifically, given the fact that a user trusts an advisor as explicitly specified by the user, if the modeled trustworthiness of the advisor is greater than a threshold, its trust value is successfully predicted to be 1. Otherwise, this is an unsuccessful prediction. For all four datasets, we also use the third metric MAE between predicted rating of each transaction (or predicted trust value for each user pair in FilmTrust, Epinions and Flixster) and the real rating of the transaction (or the actual trust relationship).

Table 3.3: Performance Comparison on the eBay Dataset

Dataset	All Data		Cold Start Buyers				Sellers Performing Inconsistently			
			$N_e \leq 2$		$N_e \leq 10$		$R_n \geq 0.01$		$R_n \geq 0.1$	
	MCC	MAE	MCC	MAE	MCC	MAE	MCC	MAE	MCC	MAE
DiffTrust	0.327	0.0648	0.007734	0.0769	0.1601	0.0699	0.3431	0.0890	0.5479	0.3010
Baseline	0.166	0.0708	-0.001773	0.0819	0.0820	0.0738	0.1722	0.0971	0.3229	0.3664
TRAVOS	0.156	0.0710	-0.002861	0.0817	0.0093	0.0784	0.1630	0.0974	0.3121	0.3679
Personalized	0.161	0.0710	-0.002861	0.0817	0.0093	0.0785	0.1684	0.0973	0.3270	0.3656
CertProp	0.270	0.0673	-0.002453	0.0812	0.0574	0.0735	0.2801	0.0922	0.4932	0.3225
Shin	0.269	0.0670	-0.002453	0.0813	0.0222	0.0732	0.2790	0.0922	0.4932	0.3225

3.2.3 Results and Discussion

Here, we present the performance of our model and the competing approaches on the four datasets, respectively. We also examine these approaches in detail by varying the time when a transaction happened and trust threshold for predicting whether a transaction is successful on the eBay dataset, and the trust threshold for predicting whether a user indicated trust on another user (advisor) on the other three datasets. Table 3.3 and Figures 3.2, 3.3 and 3.4 summarize the performance comparison of DiffTrust with other approaches in terms of MCC and MAE on eBay data in different scenarios. Table 3.4 and Figures 3.5, 3.6 and 3.7 present the results of Precision and MAE on the FilmTrust, Epinions and Flixster datasets.

More specifically, Table 3.3 shows the experimental results on (1) the whole eBay dataset with 5,331 buyers and 3,046 sellers; (2) the dataset considering only cold start buyers where $N_e \leq 2$ and $N_e \leq 10$ respectively, and N_e refers to the number of other users (advisors) with whom a buyer has shared interactions with; and (3) the dataset considering only the sellers who do not perform consistently well and $R_n \geq 0.01$ and $R_n \geq 0.1$ for sellers respectively, where R_n refers to the ratio of sellers' transactions that received negative ratings from buyers. From all these results, we can see that our DiffTrust model achieves more consistent and better performance both in MAE and MCC than other approaches. To be more specific, the performance of TRAVOS and

Figure 3.2: Performance Comparison on the eBay Dataset by Varying the Time Slots (MAE)

the personalized approach is very close to the baseline, mainly because there are only a very few shared interactions between buyers, which are needed for TRAVOS to work well. Besides, the dataset of eBay is extremely skewed and almost 99% of the transaction outcomes are positive. Thus, the personalized approach, relying on the majority of other users' ratings to evaluate advisors' trustworthiness, is more likely to underestimate the trustworthiness of advisors who provide negative ratings. In consequence, it is difficult for the personalized approach to predict the unsuccessful transactions. In our model, we adopt other buyers' direct trust evaluations on an advisor derived from their shared interactions with the advisor by considering these buyers' social proximity with respect to the current buyer. In so doing, we can assure a more accurate trust evaluation of the advisor because we are more likely to receive different trust evaluations on the advisor, especially on the advisor having ever provided negative ratings. Trust propagation approaches of Shin and CertProp are able to achieve consistently much better results than the baseline approach in most scenarios, but worse than our DiffTrust model. This is mainly because CertProp only considers the reachable buyers' direct trust on an advisor, but overlooks unreachable buyers. This may result in the loss of some important trust evaluations. Shin considers trust propagation for reachable witnesses (buyers' direct trust evaluations on the advisor), and use similarity of the buyer and unreachable witnesses'

Figure 3.3: Performance Comparison on the eBay Dataset by Varying the Time Slots (MCC)

direct trust evaluations on the same entities to model those unreachable witnesses. This setting is not very reasonable and fair for the eBay dataset. The weight of considering other buyers' direct trust evaluations on the advisor is more likely to approach 1 since most of the transactions on eBay are positive, but the weight for unreachable witnesses is more likely to be less than 1 according to the similarity metric. In contrast, in DiffTrust, we treat other buyers' direct trust evaluations towards the advisor equally by considering the social proximity. This can avoid the above mentioned "unfair" problem. It is worth noting that in the cold start cases ($N_e \leq 2$ and $N_e \leq 10$), only our model can obtain a larger positive MCC value and relatively smaller MAE value when the number of other buyers with whom a buyer has shared interactions is very sparse. In these scenarios, the performance of other approaches closely approaches that of the baseline approach. This demonstrates that our model can contribute to addressing the cold start problem, assisting new buyers to accurately evaluate advisors' trustworthiness, and thus assesses the quality of advisors' opinions to make correct decisions for buyers.

Figures 3.2, 3.3 and 3.4 illustrate the performance of different approaches by varying the time when each predicted transaction was conducted, and varying the trust threshold so that if predicted trustworthiness of a seller at specific time is higher than the threshold, the seller's transaction conducted at that time will be predicted to be successful. As shown

Figure 3.4: Performance Comparison on the eBay Dataset by Varying the Trust Threshold (MCC)

in Figures 3.2 and 3.3, we divide the time range of April 10, 2000 to June 4, 2011 into 10 time slots of equal range. We can see that the performance of all the approaches improves gradually in terms of both the MAE and MCC as time goes by. This is mainly because users gain more experience and can better model the trustworthiness of advisors. As can be seen in Figure 3.3, our model performs consistently better than other approaches, and the performance gap between our model and each other approach increases gradually as buyers gain more experience in the system. Figure 3.4 demonstrates that the best trust threshold for our model is 0.7, while 0.8 for Shin and CertProp and 0.9 for Baseline, TRAVOS and the personalized approach, respectively. Our model outperforms the other approaches for all threshold levels.

Table 3.4 depicts the results of Precision and MAE on the FilmTrust, Epinions and Flixster datasets. We can see that our model performs consistently better than other approaches. As the ratio of shared interactions between users increases (FilmTrust > Epinions > Flixster), the performance of all the approaches also improves. The performance of TRAVOS is very close to Baseline and sometimes worse than Baseline. TRAVOS works only for binary ratings. Thus, in these datasets, we first map the ratings to either positive or negative by choosing the middle of the rating scale in our experi-

Table 3.4: Performance Comparison on the FilmTrust, Epinions and Flixster Datasets

Dataset	FilmTrust		Epinions		Flixster	
	Precision	MAE	Precision	MAE	Precision	MAE
DiffTrust	0.867	0.1154	0.8163	0.2064	0.6616	0.3229
Baseline	0.711	0.2136	0.6365	0.3428	0.5193	0.4757
TRAVOS	0.682	0.3008	0.6796	0.2943	0.4884	0.464
Personalized	0.793	0.1798	0.7406	0.2266	0.5754	0.414
CertProp	0.791	0.1854	0.7327	0.2417	0.5781	0.416
Shin	0.809	0.1576	0.7413	0.2314	0.5783	0.4123

Figure 3.5: Performance Comparison on the FilmTrust by Varying the Trust Threshold

ments, which may lead to inaccurate modeling of advisors' trustworthiness according to shared interactions. For the Flixster dataset, as there are a very limited number of shared interactions between users (0.015 in Table 3.2), the performance gap between our DiffTrust model and other approaches regarding to both Precision and MAE is relatively larger than that on FilmTrust and Epinions. This is mainly because in Flixster, of all the users having shared interactions, the average number of shared interactions is relatively larger (i.e. 3.2774). That is to say, if there are any shared interactions between a user and an advisor, the user can make an accurate trust evaluation of the advisor. Through social

Figure 3.6: Performance Comparison on the Epinions by Varying the Trust Threshold

proximity, our model can maximize the use of each individual user's trust evaluation of the advisor. This demonstrates that our model, compared to other approaches, is more suitable to address the trustworthiness of advisors in the systems where users have a few shared interactions (the same trend found for eBay). It is worth noting that on these datasets, the performance of the personalized approach is much better than Baseline and TRAVOS (compared to that on the eBay dataset), implying that even in environments where users are having a relatively larger number of commonly rated items, the public information is still worth considering when addressing the trustworthiness of advisors.

Figures 3.5, 3.6 and 3.7 present the performance of different approaches by varying the trust threshold so that if a predicted trust value of advisor a by a user u is larger than the threshold, user u trusts advisor a . It shows that in general our DiffTrust model outperforms the other approaches. It consistently achieves high precision, demonstrating its effectiveness in modeling the trustworthiness of advisors.

3.3 Summary

Aiming at evaluating the quality of opinions of users (advisors) in open online environments such as e-marketplaces, we propose a novel trust model called DiffTrust, stemmed

Figure 3.7: Performance Comparison on the Flixster by Varying the Trust Threshold

from diffusion theory in Social Science, to model the trustworthiness of users. Specifically, an advisor's trust building among users is considered as a diffusion process. Her trustworthiness perceived by a specific user is influenced by four important factors including 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 the environment. Our DiffTrust model focuses on the dynamic and evolutionary characteristics of trust, and highlights that the trustworthiness of an advisor may be perceived differently by different users, depending on the environment, and is embedded with a specific context. We compare our model with a baseline approach, TRAVOS, the personalized approach, CertProp and Shin, on four real datasets of eBay, FilmTrust, Epinions and Flixster. Experimental results demonstrate that DiffTrust consistently performs better in both loosely-connected and well-connected environments. Further, it can help model the trustworthiness of advisors for users who are new to the system.

CHAPTER 4

SARC: Subjectivity Alignment for Reputation Computation

Our DiffTrust model mainly deals with the problem of advisors being dishonest in providing ratings, but overlooks the fact that users may have subjectivity difference in their opinions towards same entities. Therefore, in this chapter, we design a subjectivity alignment approach for reputation computation by specifically focusing on the the subjectivity of opinions.

A rating is a subjective evaluation of an entity by a user (referred to as an advisor) within the context of a specific interaction. Therefore, different ratings could be given for the same entities by different advisors.

Subjectivity difference may come from two sources if we analyze the scenario of an advisor providing a rating from both psychological and behavioral perspectives:

- **Intra-attribute subjectivity:** when the advisor evaluates her satisfaction level with an entity, she considers each attribute related to that entity. Although the information about each attribute is *objective*, the evaluation (i.e. satisfaction level) of the attribute value may be subjective and change from user to user. This is referred to as *intra-attribute subjectivity* in our study. For example, a product with the price of “USD1500” may be *expensive* for buyer *a*, while *not so expensive* for buyer *b*.
- **Extra-attribute subjectivity:** when the advisor assigns a satisfaction level to the entity, she may consider some attributes of the entity more heavily than others.

This is referred to as *extra-attribute subjectivity*. For example, a buyer with better economic conditions may consider a product's *quality* more heavily, while another buyer with worse economic conditions may concern more about the *price* of the product.

These two aspects together contribute to the subjectivity difference among users (or between users and advisors). Due to the subjectivity difference, it may not be effective if a user directly takes advisors' ratings towards an entity and aggregates the ratings to compute the reputation of the entity. The computed reputation values may mislead the user in selecting valuable interacting partners.

To effectively address the subjectivity difference problem, we propose a subjectivity alignment approach for reputation computation (SARC). 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 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 Bayesian learning approach to model the correlations between user a 's each rating level and the value of each objective attribute involved in the transactions. The learned correlation function, which represents user a 's *intra-attribute subjectivity*, will then be shared with the agents of other users. The agent of user a also applies a regression analysis model to learn the weight of each attribute for user a , representing her *extra-attribute subjectivity*. This information will not be shared with other users. After the learning phase, user a only needs to provide ratings for her interactions with entities, not detailed reviews.

When advisor b shares a new rating of her interaction with an entity, the agent of user a will first retrieve a rating level for each attribute of the interaction based on the shared rating and the *intra-attribute subjectivity* of advisor b shared by the agent of b . The rating levels of the attributes will then be aggregated according to user a 's *extra-attribute subjectivity* learned by the agent of a . In this way, the rating shared by advisor b is aligned

to that can be used by user a for computing the reputation of the entity.

To evaluate the performance of our SARC approach, we simulate an e-commerce environment involving a number of buyers with different subjectivity in evaluating products and a set of sellers selling products with different attribute values. In addition, buyers' subjectivity may change over time, buyers may also intentionally lie about their evaluation of products, and sellers may change the attribute values of their products. Experimental results confirm that our SARC approach provides sufficiently good performance in a general setting. It can more accurately and stably model sellers' reputation than the representative competing approaches of BLADE (Regan et al., 2006) and TRAVOS (Teacy et al., 2006). Our approach is not dramatically affected by deceptive buyers because it treats dishonest buyers as the ones with different subjectivity. It is also more robust to dynamic environments.

The rest of this chapter is organized as follows. We give an overview of our SARC approach in Section 4.1, and describe in detail how it learns users' subjectivity and aligns ratings in Section 4.2. After that, we conduct experiments to verify the effectiveness of our approach in Section 4.3. Finally, we conclude our current work in Section 4.4.

4.1 Overview of the SARC Approach

To better describe our SARC model, we choose the e-commerce environments as representative of online communities, where buyers take the role of users and advisors, and sellers are referred to as entities. Thus, in an open e-commerce environment (i.e. an e-marketplace), we denote the set of buyers by $\mathcal{B} = \{b_1, b_2, b_3, \dots\}$. The set of agents (called buying agents) equipped by corresponding buyers is denoted by $\mathcal{A} = \{a_1, a_2, a_3, \dots\}$, and the set of sellers by $\mathcal{S} = \{s_1, s_2, s_3, \dots\}$. The set of objective attributes for describing a transaction between a buyer and a seller is denoted as $\mathcal{F} = \{f_1, f_2, \dots, f_m\}$, where m represents the total number of objective attributes. Each rating provided by a buyer for a seller is from a set of predefined discrete rating levels

Table 4.1: Summary of Notations in SARC

Notations	Description
$\mathcal{B} = \{b_1, b_2, b_3, \dots\}$	Set of all buyers in the e-marketplace.
$\mathcal{S} = \{s_1, s_2, s_3, \dots\}$	Set of all sellers in the e-marketplace.
$\mathcal{A} = \{a_1, a_2, a_3, \dots\}$	Set of all agents in the e-marketplace.
$\mathcal{F} = \{f_1, f_2, \dots, f_m\}$	Set of all the objective attributes. m is its total number.
$\mathcal{L} = \{r_1, r_2, \dots, r_n\}$	Set of all the different rating levels. n is the total number.

$\mathcal{L} = \{r_1, r_2, \dots, r_n\}$, where n is the total number of different rating levels (i.e. the granularity of the rating scale). These notations are summarized in Table 4.1.

For a buyer $b_i \in \mathcal{B}$ in the marketplace, the goal of her buying agent $a_i \in \mathcal{A}$ is to accurately compute the reputation value of a target seller $s_j \in \mathcal{S}$, according to b_i 's subjectivity. In order to achieve this goal, the buying agent a_i needs to consider the ratings of other buyers (advisors) that evaluate the satisfaction levels about their past transactions with seller s_j . Due to the possible subjectivity difference between buyer b_i and the advisors, agent a_i also needs to align/convert ratings of each advisor (for example b_k) using our SARC approach.

In the SARC approach illustrated in Figure 4.1, a *Subjectivity Learner* is attached to agent a_i , by which it can model b_i 's subjectivity. More specifically, at the beginning of buyer b_i 's interactions with the system, agent a_i asks b_i to provide a rating for each of her transactions with a seller (which can be any seller in $\mathcal{S}$). Buying agent a_i also asks b_i to provide detailed review information about each transaction containing the values of the set of objective attributes in $\mathcal{F}$. Based on the provided information (rating-review pairs), agent a_i uses the *CEFs Learner* of the *Subjectivity Learner* to model a set of correlation evaluation functions (CEFs) for buyer b_i , capturing b_i 's *intra-attribute subjectivity*. Each correlation evaluation function is represented by a *Bayesian conditional probability density function* that models the correlation between each rating level of buyer b and the

Figure 4.1: Overview of the SARC Approach

value of each objective attribute. Thus, for each buyer, the total number of correlation evaluation functions is equal to $m \times n$.

The learned CEFs of buyers will be shared with each other buyer's agent. For a rating provided by the buyer (advisor) b_k , agent a_i can then derive a rating for each attribute, based on the CEFs shared by b_k 's agent a_k and those of buyer b_i 's own. Note that what is derived for an attribute is in fact a set of probability values, each of which corresponds to a rating level in $\mathcal{L}$. The rating level with the highest probability will be chosen as the rating for the attribute.

Based on the provided rating-review pairs by b_i , the *Attribute Weight Learner* of the *Subjectivity Learner* is also used by agent a_i to learn the *extra-attribute subjectivity* of buyer b_i , which is represented by a set of weights for corresponding attributes in $\mathcal{F}$. The weight of an attribute is determined by two factors: (1) the probability value of the rating derived earlier; and (2) the importance of the attribute learned using a regression analysis

model. These weights will not be shared with other buyers. Once the weights are learned, the aligned rating from that of advisor b_k can be computed as the weighted average of the derived ratings for the attributes.

In the next section, we will describe in detail how our SARC approach models CEFs based on rating-review pairs, derives a rating for each attribute, learns the weights for attributes, and computes a (aligned) rating by aggregating the derived ratings for attributes. These procedures are organized as intra-attribute subjectivity alignment and extra-attribute subjectivity alignment.

4.2 Technical Details of the SARC Approach

In this section, we describe the technical details of our SARC approach for the intra-attribute subjectivity alignment and the extra-attribute subjectivity alignment.

4.2.1 Intra-attribute Subjectivity Alignment

Given a set of rating-review pairs provided by buyer b_i , each of which is for a transaction between b_i and a seller, the rating in a pair indicates b_i 's satisfaction level about the corresponding transaction, and the review in the pair is a set of values for the attributes $\mathcal{F}$ of the transaction. Buyer b_i 's agent a_i learns the correlation evaluation functions (CEFs) of b_i , each of which is represented by a Bayesian conditional probability density function. Each CEF is the correlation between a rating level and the values of an attribute. More specifically, let us learn $\text{CEF}_{u,v}^{b_i}$, the correlation function between attribute f_u and rating level r_v for buyer b_i , where $1 \leq u \leq m$ and $1 \leq v \leq n$. Buying agent a_i first learns $p^{b_i}(r_v)$ (the probability that buyer b_i provides a rating r_v), $p^{b_i}(f_u)$ (the probability distribution of the values for attribute f_u), and $p^{b_i}(r_v \mid f_u)$ (the conditional probability of rating level r_v given the distribution of the values for attribute f_u). By applying the Bayes' Rule, agent a_i can derive $\text{CEF}_{u,v}^{b_i}$ as the conditional probability distribution of the

Figure 4.2: A Naïve Bayesian Network Model for Agent a_i of Buyer b_i to Align b_k 's Rating r^{b_k}

values for attribute f_u given rating level r_v as follows:

$$\text{CEF}_{u,v}^{b_i} = p^{b_i}(f_u | r_v) = \frac{p^{b_i}(r_v | f_u) \times p^{b_i}(f_u)}{p^{b_i}(r_v)} \quad (4.1)$$

In our SARC approach, the agents of buyers share the learned CEFs for their buyers with the agents of other buyers. Suppose that the agent a_k of a buyer b_k shares the learned CEF^{b_k} for b_k with the agent a_i of buyer b_i . For a rating r^{b_k} shared by the agent a_k of buyer b_k , agent a_i can then derive a rating level for each attribute in $\mathcal{F}$. We use a Naïve Bayesian Network model to learn the mapping/alignment from r^{b_k} of buyer b_k to the ratings of b_i for the attributes, as illustrated in Figure 4.2. Although in this model we assume that the attributes are independent given the ratings of buyers, in Section 4.2.2, we will learn the relative weights of the attributes to capture the dependency among the attributes.

Let us take any $f_u \in \mathcal{F}$ as an example attribute to show how agent a_i derives a rating for attribute f_u . To do so, agent a_i first estimates the conditional probability of a rating level in $\mathcal{L}$ for attribute f_u , given rating r^{b_k} provided by buyer b_k . Take any rating level r_v as an example, agent a_i computes $p^{b_i}(r_{v,f_u} | r^{b_k})$, the conditional probability that buyer b_i

will assign the rating level r_{v,f_u} to attribute f_u given the rating r^{b_k} of buyer b_k , as follows:

$$\begin{aligned} p^{b_i}(r_{v,f_u}|r^{b_k}) &= \frac{p^{b_i}(r_v | f_u, r^{b_k}) \times p^{b_k}(f_u | r^{b_k})}{p^{b_i}(f_u | r_v, r^{b_k})} \\ &= \frac{p^{b_i}(r_v | f_u) \times p^{b_k}(f_u | r^{b_k})}{p^{b_i}(f_u | r_v)} \end{aligned} \quad (4.2)$$

where $p^{b_k}(f_u | r^{b_k})$ is learned by agent a_k of buyer b_k using Equation 4.1 and shared by agent a_k to agent a_i , $p^{b_i}(f_u | r_v)$ is learned by agent a_i itself using Equation 4.1, and $p^{b_i}(r_v | f_u)$ is obtained by agent a_i from the rating-review pairs provided by its buyer b_i . In Equation 4.2, $p^{b_i}(r_v | f_u, r^{b_k})$ is equivalent to $p^{b_i}(r_v | f_u)$ and $p^{b_i}(f_u | r_v, r^{b_k})$ is equivalent to $p^{b_i}(f_u | r_v)$ because buyer b_i provides ratings to corresponding attributes regardless of buyer b_k 's ratings. In other words, buyers evaluate transactions independently.

For attribute f_u , agent a_i learns the conditional probability of each rating level $r_v \in \mathcal{L}$ according to Equation 4.2. The aligned rating of attribute f_u for buyer b_i on the basis of buyer b_k 's rating is then determined as the rating level with the highest probability value, as follows:

$$r_{u,k}^{b_i} = \arg \max_{r_v \in \mathcal{L}} p^{b_i}(r_{v,f_u}|r^{b_k}) \quad (4.3)$$

The aligned ratings for other attributes in $\mathcal{F}$ can also be determined in the same way according to Equations 4.2 and 4.3.

For example, assume that there are five rating levels 1, 2, 3, 4 and 5, and three objective attributes f_1 , f_2 and f_3 . Buyer b_k provides a rating level 3 to buyer b_i . Through the intra-attribute subjectivity alignment, agent a_i gets that, for this experience, the rating level 3 of b_k means that b_i will evaluate attribute f_1 as rating level 1, attribute f_2 as 3, and attribute f_3 as 4, respectively.

4.2.2 Extra-attribute Subjectivity Alignment

After the ratings of the attributes are obtained, agent a_i of buyer b_i then aggregates the ratings to represent an aligned rating of the rating r^{b_k} shared by buyer b_k . To do this, a_i needs to first determine a weight for each attribute in $\mathcal{F}$, because buyer b_i may be more concerned about one attribute over another.

The weight of an attribute f_u is determined by two factors. One factor is the confidence C_u about the rating $r_{u,k}^{b_i}$ derived for the attribute f_u using Equations 4.2 and 4.3. The confidence can be represented as the conditional probability value of the derived rating, $p^{b_i}(r_{u,k}^{b_i} | r^{b_k})$ estimated using Equation 4.2. A larger probability value means that it is more probable that the derived rating for attribute f_u should be $r_{u,k}^{b_i}$ according to buyer b_k 's rating and the subjectivity of buyers b_i and b_k . In other words, the larger the probability is, the more reliable the derived rating $r_{u,k}^{b_i}$ is. Thus, we have:

$$C_u = p^{b_i}(r_{u,k}^{b_i} | r^{b_k}) \quad (4.4)$$

Another factor to determine the weight for attribute f_u is the importance I_u of f_u in buyer b_i 's view. The importance I_u can be modeled as the coefficient of attribute f_u by a regression analysis model, based on the rating-review pairs provided by b_i . More specifically, given the rating-review pairs, we compute the coefficients for attributes by minimizing the aggregated difference between the true ratings in the rating-review pairs of b_i and the ratings, each of which is predicted for a review by the following equation:

$$r_0^{b_i} = I_0 + \sum_{u=1}^m I_u \times V_{f_u} + \varepsilon \quad (4.5)$$

where $r_0^{b_i}$ is the predicted rating for a review, V_{f_u} is the value of f_u in the review, I_0 is a constant, and ε is residual. So, the coefficients $I = [I_0, I_1, \dots, I_m]$ can be computed by:

$$I' = (X'X)^{-1}X'Y \quad (4.6)$$

where if there are c rating-review pairs for buyer b_i in total,

$$X = \begin{bmatrix} 1 & f_{11} & \cdots & f_{m1} \\ 1 & f_{12} & \cdots & f_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & f_{1c} & \cdots & f_{mc} \end{bmatrix}, \quad Y = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_c \end{bmatrix} \quad (4.7)$$

After the weight (confidence and importance) of each attribute is determined, the aligned rating $r_k^{b_i}$ can be computed as the weighted average of the ratings for attributes derived using Equations 4.2 and 4.3, as follows:

$$r_k^{b_i} = \frac{\sum_{u=1}^m r_{u,k}^{b_i} \times C_u \times I_u}{\sum_{u=1}^m C_u \times I_u} \quad (4.8)$$

Following the example in the previous section, based on b_i 's past experience, agent a_i obtains b_i 's weights of f_1 , f_2 and f_3 as 0.1, 0.2 and 0.9, respectively. In this case, the final rating for b_i from b_k 's rating level 3 is computed as: $(0.1 \times 1 + 0.2 \times 3 + 0.9 \times 4) / (0.1 + 0.2 + 0.9) = 3.58 \approx 4$.

After aligning all ratings shared by all buyers (advisors), the reputation value of seller s_j in the view of b_i can be computed as, for example, the average of the aligned ratings.

4.3 Experiments

In this section, we carry out experiments to evaluate the performance of our SARC approach and compare it with some representative competing approaches.

4.3.1 Experimental Environment

We simulate an e-commerce environment involving 50 sellers and 200 buyers¹⁰ In our simulations, sellers may provide different products. Their products under the category of computer are represented by five objective attributes, namely, *Price*, *Speed of CPU*, *Processor Type*, *Graphics Card Type*, and *Hard Drive Size* with ranges presented in Table 4.2. For each seller, the values of the five attributes of her products are randomly chosen within these ranges.

Buyers may have different subjectivity in evaluating their transactions with (the products of) sellers. We simulate both buyers' intra-attribute subjectivity and extra-attribute subjectivity. To be specific, we assume that a buyer's rating for a transaction with a seller is derived as follows. First, the buyer evaluates each objective attribute according to a specific intrinsic (taste) function. In our experiments, buyers' intra-attribute subjectivity is simulated as approximate *Gaussian Distribution*. That is, for each attribute, the probability of each rating level given by a buyer is in the form of normal distribution. Second, the buyer places random weights (in the domain of $[0,1]$) on different attributes, and computes the weighted average of her evaluations on attributes as a single rating for the transaction. Since buyers can only give ratings under the predefined rating scale in reality, the simulated rating is chosen from the predefined rating scale that is the closest to the weighted average.

In the experiments, besides our SARC approach, we implement a baseline approach without subjectivity alignment, which computes the reputation of sellers by directly averaging the ratings collected from other buyers for the sellers. We also choose to implement the TRAVOS approach (Teacy et al., 2006), which is a representative approach in the set of filtering approaches (see Chapter 2 for details). The BLADE approach (Regan et al., 2006) is chosen instead of the approach of Koster et al. (2010) because the two approaches are very similar and the approach of Koster et al. (2010) is complicated to

¹⁰We use simulation, instead of the real datasets mentioned in our thesis, mainly because the attribute (objective or not) information about an entity is not available across all these datasets.

Table 4.2: Product Attributes and Value Ranges

Dimension	Type	Ranges
Price	Double	100-10,000
Speed of CPU	Double	1-10 GHZ
Processor Type	Char	5 types
Graphics Card Type	Char	2 types
Hard Drive Size	Integer	40-1000GB

implement.

We compare the performance of these approaches with our approach in computing the reputation of sellers. The performance of an approach is measured as the mean absolute error (MAE) between the reputation of sellers computed for each buyer using the approach, and the reputation of sellers using the ratings according to each buyer's own subjectivity (representing the ground truth about the reputation of sellers with respect to the buyer).

4.3.2 Experimental Parameters

To simulate real-world e-commerce environments, we set several important parameters for our simulations, including information availability, dynamic behavior of sellers, dynamic subjectivity of buyers, ratio of liars (dishonest buyers), and granularity of rating scale.

Information availability refers to the amount of available information required by different approaches for subjectivity alignment. Two types of information are needed by our approach. One type of information is the detailed reviews describing the objective attributes of transactions between buyers and sellers. This information is used by our approach to model the correlation evaluation functions (CEFs) and the importance of the attributes for buyers. In the experiments, we vary the *number of detailed reviews* (N_r) to see how the performance of our approach is affected by this parameter. Another type of information contributing to our approach is the number of objective attributes. In reality,

some attributes (e.g. *appearance*) may not be objective. The total number of objective attributes in our simulations may thus be less than 5. In the experiments, we vary the *ratio of objective attributes* (R_{obj}) to be 0%, 20%, 40%, 60%, 80% and 100%, to see how much the performance of our approach will be affected. One type of information required by the BLADE approach is shared interactions where buyers and advisors have interacted with some common sellers. We vary the *ratio of shared interactions* (R_i) to see how this parameter affects the performance of BLADE.

We also set the parameter P_{seller} to capture the *dynamic behavior of sellers*. In real-world e-commerce environments, sellers may change their behavior over time. For example, they may provide products of high quality at first, but those of low quality after earning enough reputation. In our experiments, dynamic behavior of sellers is simulated by changing the quality of their products (i.e. the values of a subset of the objective attributes in Table 4.2).

Buyers in the marketplace may also adjust their subjectivity over time. Dynamic subjectivity of buyers (P_{buyer}) is captured in their rating procedure by adjusting “*intra-attribute subjectivity*”, or “*extra-attribute subjectivity*”, or both.

Ratio of liars (R_{liar}) is adopted in our simulations to reflect the deception problem in real e-marketplaces where some buyers may lie about their experience with sellers. Following the work of (Yu and Singh, 2003a; Whitby et al., 2004; Teacy et al., 2006; Şensoy et al., 2009), we also simulate the complementary lying behavior where if a true rating to a seller is r in the scale of $[0, 1]$, the liar will modify the rating as $1 - r$.

Granularity of rating scale (G_{scale}) refers to the number of rating levels. It may be different for different reputation systems. In our experiments, we will study the effect of the granularity of rating scale by varying G_{scale} from 2 to 10.

We vary the values of the above parameters to simulate basic, deceptive and dynamic environments, respectively.

4.3.3 Experimental Results

Here, we present the performance of our approach and the competing approaches in different simulated environments. First, we investigate how the approaches perform in the basic environment without deception or variation on buyers' subjectivity and sellers' behavior. Various experiments are conducted by varying the other related parameters that may influence the performance of the approaches. Second, we examine the performance of these approaches in the deceptive environment. Third, to consider the dynamic scenarios, we inspect these approaches by adopting the dynamic behavior of sellers and dynamic subjectivity of buyers respectively.

Basic Environment

We first simulate a basic environment without any variation of the parameters (i.e. $R_{liar} = 0$, $P_{seller} = 0$, $P_{Buyer} = 0$), and compare the performance of our approach and that of the three competing approaches, including the baseline approach, TRAVOS and BLADE. We compute their mean absolute error (MAE) values for computing the reputation of sellers in different epoches. In each epoch, each buyer interacts with one seller in the marketplace. From the results shown in Figure 4.3, we can see that our approach performs consistently the best no matter whether buyers have more or less experience with sellers. Because both TRAVOS and BLADE require shared interactions, their performance is limited. Both TRAVOS and BLADE perform slightly better than the baseline approach. The performance difference between the different approaches is reduced when buyers have more experience with sellers in the marketplace.

Based on the basic environment, we then vary some parameters to examine their effects. We first examine how the ratio of objective attributes R_{obj} affects our SARC approach. We vary R_{obj} from 0% to 100% for our SARC approach, while keep R_{obj} to be 100% for BLADE. As shown in Figure 4.4, SARC performs slightly worse than BLADE when there are no objective attributes. However, it performs better than BLADE when there are more than 20% of objective attributes. The performance of SARC consistently

Figure 4.3: Performance Comparison in the Basic Environment Where $R_{liar} = 0$, $P_{seller} = 0$, $P_{Buyer} = 0$

Figure 4.4: Performance When Varying the Ratio of Objective Attributes

increases as the ratio of objective attributes increases. But, the increment becomes smaller when $R_{obj} \geq 20\%$.

The finer the granularity of rating scale (G_{scale}) is, the easier to learn buyers' subjectivity because buyers' subjectivity can be better captured by the larger granularity of rating scale. This trend is verified by our experiment. In Figure 4.5, we plot the MAE results of the four approaches when varying G_{scale} from 2 to 10. The figure shows that the performance of SARC is significantly greater than the baseline approach, TRAVOS and BLADE. On average, the performance of SARC improves as G_{scale} increases.

Figure 4.5: Performance When Varying the Granularity of Rating Scale

Figure 4.6: Performance When Varying the Number of Detailed Reviews

We also vary the number of detailed reviews (N_r) provided by buyers from 1 to 30. We try to determine a reasonable N_r for SARC. As shown in Figure 4.6, when N_r increases from 1 to 5, the performance of SARC increases significantly. While N_r is larger than 5, as the increase of N_r , the performance of SARC also increases, but to a much smaller degree. This is simply because SARC requires only a few detailed reviews to learn buyers' subjectivity well. After that, any additional information leads to only small improvement, and so, we can choose 6 as the acceptable minimum N_r . SARC performs better than the baseline approach and BLADE in all the cases for N_r .

As discussed in Chapter 2, the BLADE model requires shared interactions in order

Figure 4.7: Varying Ratio of Shared Interactions

to learn buyers' subjectivity. However, in real e-marketplaces, shared interactions are generally very sparse. In this experiment, we fix the number of past interactions for each buyer, but vary the ratio of shared interactions (R_i) from 0% to 100%. For each ratio value, MAE is computed as the average of five repeated runs. Figure 4.7 indicates that BLADE performs significantly worse than SARC when R_i is in the range from 0% to 30%. The performance of BLADE increases with the increase of R_i .

Deceptive Environment

In this experiment, we examine the effect of deception (buyers lying about their past experience) on different approaches. We vary the ratio of liars (R_{liar}) from 0% to 100%, and plot the MAE results of different approaches in Figure 4.8. We can see that the performance of TRAVOS does not decrease much as R_{liar} increases. Our SARC performs much better than the other three models for any R_{liar} . It is not dramatically affected by lying buyers because SARC learns a buyer's subjectivity from the buyer's own past experience and treats lying buyers as the buyers with different subjectivity. When R_{liar} is larger than 0.5, BLADE performs worse than TRAVOS, but consistently better than the baseline approach. Note that in the environment where most buyers are liars, the performance of other models are not so bad. This is mainly because buyers have different

Figure 4.8: Performance When Varying Ratio of Lying Buyers

subjectivity in our simulations. The effect of buyers' lying behavior may be reduced by the subjectivity difference among buyers, and vice versa.

Dynamic Environment

In this experiment, we simulate the environment where sellers may change the quality of their provided products in their transactions with buyers. We define a predefined parameter, P_{seller} , to represent the probability that each seller may vary the values of the five attributes of her provided products. We assume that sellers only change their behavior once in the marketplace. Once their behavior is changed, they will keep the behavior. P_{seller} is ranged from 0 to 1 and increased by 0.05 in our experiment. The MAE results for SARC and other three approaches are plotted in Figure 4.9, which demonstrates that the performance of SARC is not sensitive to the dynamic behavior of sellers, and it performs almost consistently in all cases, while the performance of Baseline, TRAVOS and BLADE gets worse as P_{seller} increases. The main reason is that SARC models the rating behavior (subjectivity) of each buyer from the buyer's own experience, which is independent of sellers' behavior change. For TRAVOS and BLADE, they rely on past shared interactions between the buyer and advisors, and these shared interactions may not be appropriate to be source information used for aligning the buyer's subjectivity due

Figure 4.9: Performance When Varying Probability of Sellers' Changing Behavior

to the possible behavior change of sellers in the shared interactions. For example, for a buyer and an advisor with the same subjectivity, if they interact with a seller in different time periods where the seller has changed behavior, TRAVOS may incorrectly treat the advisor as a liar and BLADE may incorrectly conclude that the buyer and the advisor have different subjectivity.

In a marketplace, buyers may also change or adjust their subjectivity after several interactions with sellers. In this experiment, we assume that buyers will change their subjectivity with a certain predefined probability, P_{buyer} . As in the previous experiment, buyers only change their subjectivity once in the marketplace and then keep their changed subjectivity in the following interactions with sellers. Figure 4.10 shows that the performance of SARC is not affected by buyers' dynamic subjectivity. In SARC, buying agents can update the learned subjectivity of buyers by acquiring their buyers' own recent experience, which provides flexibility to deal with buyers changing subjectivity. The performance of BLADE becomes almost equivalent to that of Baseline as P_{buyer} increases, and is consistently lower than SARC. In BLADE, once a buyer's subjectivity is changed, her buying agent cannot align ratings from advisors effectively because new shared interactions between the buyer and advisors are needed. TRAVOS performs worse than Baseline as P_{buyer} increases because the learned results of advisors become

Figure 4.10: Performance When Varying Probability of Buyers' Changing Subjectivity

misleading after they change subjectivity.

4.3.4 Discussions on Experimental Results

In this section, we summarize our experimental results. In Section 4.3.3, we compare the performance of our SARC approach in addressing the subjectivity difference problem for reputation computation with other three representative models: Baseline, TRAVOS and BLADE. In the environments without deception, seller dynamic behavior or buyer dynamic subjectivity, SARC can more accurately model sellers' reputation than BLADE, TRAVOS and Baseline. In this environment, we also test the influence of some parameters including the ratio of objective attributes (R_{obj}), the number of detailed reviews (N_r), and the granularity of rating scale (G_{scale}), all of which may affect the performance of our approach. We find that in different settings, SARC still has better performance than BLADE. Further, when $R_{obj} \geq 40\%$ or $N_r \geq 6$, SARC can attain relatively stable and acceptable performance, indicating that the performance of SARC is relatively insensitive to the amount of available information, and it can achieve good performance with limited information. We also examine the effort of the ratio of shared interactions (R_i) on the BLADE model. The result demonstrates that the performance of SARC is significantly greater than BLADE when the ratio of shared interactions is in the range $[0, 0.3]$.

In the environments where some buyers may intentionally lie about their past experience with sellers, our SARC approach still performs much better than the other approaches. It is not dramatically affected by buyers' deception because it treats deceptive buyers as the ones with different subjectivity, and aligns the ratings from them effectively. Among the four approaches, TRAVOS is the most robust to the deceptive environment. BLADE is worse than TRAVOS when the percentage of liars is greater than 0.5.

In the dynamic environment where sellers may change their provided products, SARC performs consistently and is independent of the change of P_{seller} . The performance of other three approaches gets worse as sellers are more likely to change their behavior. In the environment where buyers may vary their subjectivity during a certain period of their interactions with sellers, our experiment shows that SARC continues to perform positively, while the performance of BLADE gets closer to the baseline approach. The TRAVOS model performs worse than the baseline approach when P_{buyer} increases.

In summary, the experimental results confirm that our SARC approach has better performance than the other competing approaches in different simulated environments. It is also not dramatically influenced by deceptive buyers, and more robust to the dynamic environments. Although SARC requires objective attributes and detailed reviews, the requirements are actually not very restrictive and can be easily satisfied in real-world e-commerce environments.

4.4 Summary

In this chapter, we propose a subjectivity alignment approach for reputation computation, SARC, to address the subjectivity difference problem. In SARC, users' subjectivity is learned based on the ratings and detailed reviews they provide about the objective attributes of their interactions with entities. More specifically, SARC separately learns the *intra-attribute subjectivity* and *extra-attribute subjectivity* of users (or advisors). Their

intra-attribute subjectivity is modeled using Bayesian learning. Their *extra-attribute subjectivity* is learned using a regression analysis model. We also conduct various experiments in a simulated e-commerce environment to compare the performance of our approach with that of other three competing models, including the baseline approach, TRAVOS and BLADE. Experimental results demonstrate that: (1) SARC performs better than the other three approaches, and can more accurately and stably model sellers' reputation; (2) SARC is capable of coping with environments with deception and dynamic buyer and seller behavior; and (3) the requirement of detailed reviews and objective attributes is not very restrictive.

CHAPTER 5

PGTM: Probabilistic Graphical Trust Model

Our DiffTrust model, inspired by diffusion theory in Social Science, mainly deals with advisors' dishonesty problem when modeling advisor's trustworthiness, while the proposed SARC model addresses two kinds of subjectivity of users' opinion evaluation. In this chapter and the next chapter, we will present two novel models that simultaneously consider both the dishonesty and subjectivity difference aspects, and can distinguish between them when evaluating the quality of opinions.

To recap, as indicated in Chapter 1, we identify two dominant reasons leading to diversity of advisors' opinion quality: (1) some advisors might be dishonest due to their intrinsic nature such as a 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 contradict with their real experience; and (2) advisors might be honest, but are *subjectively different* (Fang et al., 2012b) 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.

Some existing trust models only consider either advisors' dishonesty (Zhang and Cohen, 2007) or subjectivity difference (Fang et al., 2012b), 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 of 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 chapter, we propose a novel probabilistic graphical trust model (PGTM) that explicitly distinguishes between 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.

The rest of the chapter is organized as follows. In Section 5.1, we firstly carefully present each step of our PGTM model, including the settings of parameters and the model generative process, model inference and parameter estimation, and the trust prediction based on the learned model. Next, we conduct experiments to verify the effectiveness of our model in trust modeling in Section 5.2. Finally, we summarize this work in Section 5.3.

5.1 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 5.1. 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 of advisor

Figure 5.1: The Conceptual Framework of Trust

trustworthiness (Mayer et al., 1995). In reality, users' different backgrounds will conduce to their different initial trust towards others. Different users may have different subjectivity in evaluating the same entities. Even if an advisor is honest, a user may have different experiences 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 differences between the user and the advisor, as well as that between the advisor and other advisors (users) can have a 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.

5.1.1 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

Table 5.1: Summary of Notations in PGMT

Notations	Description
u	A user.
a	An advisor.
y_u	u 's trust propensity.
c_a	a 's competency.
b_a	a 's benevolence.
i_a	a 's integrity.
$s_{u,a}$	a 's subjectivity difference with u .
r	Rating difference towards the same entity between every user and an advisor.
R	Rating difference between an advisor and the average of all other users for the same entity.

trust links are available. We thus design two graphical models, shown in Figures 5.2(a) and 5.2(b), for the two types of communities, respectively. We use graphical models mainly because they can fully interpret our conceptual framework, and seamlessly merge supervised and unsupervised learning (labeled and unlabeled relationships) (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 full trust and 0 full distrust¹¹. 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 the range $[0, 1]$. Specifically, the expected *trust propensity* y_u of 1 represents complete propensity to trust, while that of 0 no propensity. These notations are summarized in Table 5.1.

Competence (c) indicates whether an advisor has the ability to provide reliable ratings

¹¹In 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.

(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. A higher benevolence of the advisor with respect to users will lead to a smaller rating difference between every user and the advisor. Through the chains in the graphical models (see Figure 5.2, 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) represents the degree that a person follows rules in an organization (McKnight and Chervany, 2001). Then, it can be inferred that integrity affects rating differences (R) between the advisor and the average of all other users for the same entity. Higher integrity implies smaller rating difference. Similar to benevolence, we model the relationship of integrity with advisor trustworthiness (see Figure 5.2, 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 capture 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 5.2(b) is a special case for Figure 5.2(a), of which the expected expressiveness e_u always equals 0.

Figure 5.2 shows dependencies among variables and parameters. The directed arrows represent dependency relationships. Variables in shaded circles (e.g. $r_{u,a,k}$ and $o_{u,a}$)

Figure 5.2: Graphical Model: (a) with (b) without Trust Links

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 5.2(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 $y_u \sim \text{Beta}(\psi)$.

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 $c_a \sim \text{Beta}(\alpha)$, $b_{a|t} \sim \text{Beta}(\varphi)$, and $i_{a|t} \sim \text{Beta}(\theta)$.

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), using Bernoulli distributions based on u ’s propensity y_u and a ’s competence c_a , as $t_{u,a} \sim \text{Bern}(y_u) \cdot \text{Bern}(c_a)$; **[3.2]** u samples distribution parameters: expressiveness $e_{u|t}$ and subjectivity difference $s_{u,a|t}$ using Beta distribution with symmetric parameters, as $e_{u|t} \sim \text{Beta}(\varepsilon)$ and $s_{u,a|t} \sim \text{Beta}(\beta)$; **[3.3]** u generates observability of link $o_{u,a}$ using Bernoulli distribution based on u ’s expressiveness e_u and trust of the link $t_{u,a}$,

as $o_{u,a} \sim \text{Bern}(e_{u|t})$; [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, $R_{U-a,a,k}$, using Binomial distributions, as $r_{u,a,k} \sim \text{Bin}(m, 1-b_{a|t}) \cdot \text{Bin}(m, s_{u,a|t})$ and $R_{U-a,a,k} \sim \text{Bin}(m, 1-i_{a|t})$, where $R_{U-a,a,k}$ and $r_{u,a,k}$ are in $[0, m]$.

We use a 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 a 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.

5.1.2 Model Inference and Parameter Estimation

We use Gibbs sampling (Casella and George, 1992) to infer our two models (Figures 5.2(a) and 5.2(b)), and then update the 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 the 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 5.2, for model inference.

When trust links are partially observable, as shown in Figure 5.2(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:

$$\begin{aligned}
 &P(t_{u,a} = t | t_{u,-a}, o, r, R, \alpha, \varphi, \theta, \beta, \psi, \varepsilon) \\
 &\propto P(t_{u,a} = t | t_{u,-a}, \alpha, \psi) P(R_{U-a,a} | t_{u,a} = t, R_{U-a,-a}, \theta) \\
 &P(o_{u,a} | t_{u,a} = t, o_{u,-a}, \varepsilon) P(r_{u,a} | t_{u,a} = t, r_{u,-a}, \varphi, \beta)
 \end{aligned} \tag{5.1}$$

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 advisor other than a .

We then evaluate the above 4 components independently:

$$P(o_{u,a} | t_{u,a} = t, o_{u,-a}, \varepsilon) = \frac{n(o_u^1 | t) + \varepsilon}{n(o_u^0 | t) + n(o_u^1 | t) + 2\varepsilon} \tag{5.2}$$

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 $o_{u,-a}$.

$$P(t_{u,a} = t | t_{u,-a}, \alpha, \psi) = \left[\frac{n(t_a^t) + \alpha}{n(t_a^0) + n(t_a^1) + 2\alpha} \right] \left[\frac{n(t_u^t) + \psi}{n(t_u^0) + n(t_u^1) + 2\psi} \right] \tag{5.3}$$

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 .

$$\begin{aligned}
P(r_{u,a}|t_{u,a} = t, r_{u,-a}, \varphi, \beta) = & \\
& \left[\Gamma \left(\sum_{r=0}^m m.n(r_a^r|t) + 2\varphi \right) \Gamma \left(\sum_{r=0}^m r [n(r_a^r|t) + n(r_{u,a}^r|t)] + \varphi \right) \right. \\
& \left. \Gamma \left(\sum_{r=0}^m (m-r) [n(r_a^r|t) + n(r_{u,a}^r|t)] + \varphi \right) \right] \left[\Gamma \left(\sum_{r=0}^m r.n(r_a^r|t) + \varphi \right) \right. \\
& \left. \Gamma \left(\sum_{r=0}^m (m-r)n(r_a^r|t) + \varphi \right) \left(m \sum_{r=0}^m [n(r_a^r|t) + n(r_{u,a}^r|t)] + 2\varphi \right) \right]^{-1} \quad (5.4) \\
& \left[\Gamma \left(\sum_{r=0}^m m.n(r_{u,a}^r|t) + 2\beta \right) \Gamma \left(\sum_{r=0}^m 2r [n(r_{u,a}^r|t)] + \beta \right) \right. \\
& \left. \Gamma \left(\sum_{r=0}^m 2(m-r) \cdot [n(r_{u,a}^r|t)] + \beta \right) \right] \left[\Gamma \left(\sum_{r=0}^m r.n(r_{u,a}^r|t) + \beta \right) \right. \\
& \left. \Gamma \left(\sum_{r=0}^m (m-r).n(r_{u,a}^r|t) + \beta \right) \left(2m \sum_{r=0}^m n(r_{u,a}^r|t) + 2\beta \right) \right]^{-1}
\end{aligned}$$

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.

$$\begin{aligned}
P(R_{U-a,a}|t_{u,a} = t, R_{U-a,-a}, \theta) & \\
= & \left[\Gamma \left(\sum_{r=0}^m m.n(R_a^r|t) + 2\theta \right) \Gamma \left(\sum_{r=0}^m r [n(R_a^r|t) + n(R_{U-a,a}^r|t)] + \theta \right) \right. \\
& \left. \Gamma \left(\sum_{r=0}^m (m-r) \cdot [n(R_a^r|t) + n(R_{U-a,a}^r|t)] + \theta \right) \right] \quad (5.5) \\
& \left[\Gamma \left(\sum_{r=0}^m r.n(R_a^r|t) + \theta \right) \Gamma \left(\sum_{r=0}^m (m-r).n(R_a^r|t) + \theta \right) \right. \\
& \left. \Gamma \left(m \sum_{r=0}^m [n(R_a^r|t) + n(R_{U-a,a}^r|t)] + 2\theta \right) \right]^{-1}
\end{aligned}$$

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 .

When trust links are not observable, as shown in Figure 5.2(b). The corresponding sampling process is:

$$\begin{aligned} P(t_{u,a} = t | r, R, \alpha, \varphi, \theta, \beta, \psi) &\propto P(t_{u,a} = t | \alpha, \psi) \\ P(r_{u,a} | t_{u,a} = t, r_{u,-a}, \varphi, \beta) &P(R_{U-a,a} | t_{u,a} = t, R_{U-a,-a}, \theta) \end{aligned} \quad (5.6)$$

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:

$$\begin{aligned} P(y_u | t_u, \psi, \alpha) &\sim \text{Beta}(\psi + n(t_u^1), \psi + n(t_u^0)) \\ P(c_a | t_a, \psi, \alpha) &\sim \text{Beta}(\alpha + n(t_a^1), \alpha + n(t_a^0)) \\ P(e_{u|t} | t_u, o_u, \varepsilon) &\sim \text{Beta}(\varepsilon + n(o_u^1 | t), \varepsilon + n(o_u^0 | t)) \\ P(s_{u,a|t} | t_a, r_a, b_{a|t}, \beta) &\sim \\ &\text{Beta}\left(\sum_{r=0}^m r n(r_{u,a}^r | t) + \beta, \sum_{r=0}^m (m-r) n(r_{u,a}^r | t) + \beta\right) \\ P(i_{a|t} | t_a, R_a, \theta) &\sim \text{Beta}\left(\sum_{r=0}^m \left[r(1-t)n(R_a^r | t) + t(m-r)n(R_a^r | t) \right] \right. \\ &\quad \left. + \theta, \sum_{r=0}^m \left[rtn(R_a^r | t) + (1-t)(m-r)n(R_a^r | t) \right] + \theta \right) \end{aligned} \quad (5.7)$$

$$\begin{aligned} P(b_{a|t} | t_a, r_a, s_a, \varphi) &\sim \text{Beta}\left(\sum_{r=0}^m \left[r(1-t)n(r_a^r | t) + t(m-r)n(r_a^r | t) \right] \right. \\ &\quad \left. + \varphi, \sum_{r=0}^m \left[rtn(r_a^r | t) + (1-t)(m-r)n(r_a^r | t) \right] + \varphi \right) \end{aligned} \quad (5.8)$$

5.1.3 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 5.2. 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 $t_{u,a} = t$ as follows:

$$\begin{aligned}
 &P(t_{u,a} = t | o_{u,a}, r_{u,a}, R_{U-a,a}, c_a, b_a, i_a, s_{u,a}, y_u, e_u) \\
 &\propto P(t_{u,a} = t | y_u, c_a) P(R_{U-a,a} | t_{u,a} = t, i_a) \\
 &P(o_{u,a} | t_{u,a} = t, e_u) P(r_{u,a} | t_{u,a} = t, b_a, s_{u,a})
 \end{aligned} \tag{5.9}$$

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

5.2 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.

5.2.1 Data Description

Each of our datasets consist 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 rated entities between users. To obtain our data, we process three available datasets, including

Table 5.2: Statistical Information about the Three Datasets

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

FilmTrust, Flixster and Epinions. Specifically, we use the original FilmTrust dataset, and randomly sample a small portion of original Flixster dataset. For the 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 5.2.

5.2.2 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 Chapter 2. For the 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.

5.2.3 Evaluation Metrics

To measure the performance of models, we use the commonly used metrics, including *precision*, *recall*, *f-value*, and *MAE*. $\text{Precision} = \frac{t_p}{t_p + f_p}$, $\text{recall} = \frac{t_p}{t_p + f_n}$, and $\text{f-value} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$, 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.

5.2.4 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, then u trusts a .

Model Effectiveness

In Figures 5.3, 5.4, 5.5 and 5.6, we analyze the effectiveness of each latent variable in our model by showing the performance of the 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

Figure 5.3: Model Effectiveness on the Epinions Dataset by Varying the Trust Threshold (F-Value)

Figure 5.4: Model Effectiveness on the Epinions Dataset by Varying the Trust Threshold (Precision)

($\text{PGTM}(y, c, b, s)$), with all but *expressiveness* latent variable ($\text{PGTM}(y, c, b, s, i)$) and the complete model ($\text{PGTM}(\text{full})$) which includes all latent variables.

As can be seen, $\text{PGTM}(y, c)$ performs better than $\text{PGTM}(y)$ when the trust threshold is larger than 0.5. This might be due to the fact that $\text{PGTM}(y, c)$ rather than $\text{PGTM}(y)$ can enable risk-averse users (with higher trust thresholds) to adjust their propensity to (dis)trust according to advisors' competence through interactions. Next, the superiority of $\text{PGTM}(y, c, b, s)$ over $\text{PGTM}(y, c)$ becomes more salient when the number of commonly

Figure 5.5: Model Effectiveness on the Flixster Dataset by Varying the Trust Threshold

Figure 5.6: Model Effectiveness on the FilmTrust Dataset by Varying the Trust Threshold

rated entities between users and advisors increases (**Epinions**>**FilmTrust**>**Flixster** as shown in Table 5.2). 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 differences between users and advisors. With more commonly rated items, we can model rating differences more accurately.

Moreover, $\text{PGTM}(y, c, b, s, i)$ outperforms $\text{PGTM}(y, c, b, s)$ 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 $\text{PGTM}(\text{full})$ performs best because instead of using

Figure 5.7: Model Effectiveness (Subjectivity vs. Dishonesty) (F-Value)

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 Figures 5.7 and 5.8, where *subjectivity* denotes the model with only *propensity* and *subjectivity difference* latent variables, while *dishonesty* denotes the model with *propensity* and *competence*, *benevolence*, and *integrity* latent variables. It can be observed that $\text{PGTM}(y, c, b, s, i)$ outperforms the *subjectivity* model and *dishonesty* model, demonstrating that the subjectivity difference factor, which is paid less attention before, in addition to the dishonesty factor, improves the performance of our model. We also find that 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 the subjectivity-related latent variable (i.e. subjectivity difference) when the trust threshold is larger than 0.5.

Model Comparison

Table 5.3 and Figures 5.9, 5.10, 5.11 and 5.12 show the performance comparisons between our approach ($\text{PGTM}(y, c, b, s, i)$) and other approaches on three datasets. In order to

Figure 5.8: Model Effectiveness (Subjectivity vs. Dishonesty) (Precision)

Table 5.3: Performance Comparison on the FilmTrust, Epinions and Flixster Datasets

Dataset Approach	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

make fair comparisons, we assume that all trust links are not observable in our model (i.e. we adopt the model in Figure 5.2 (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 the Epinions dataset. Second, as shown in Table 5.3, 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 with Epinions and FilmTrust, but worse than that with Flixster. This is mainly because there are fewer commonly rated entities in the Flixster dataset than in the other two datasets (see Table 5.2). Without sufficient commonly rated items, Prob-Cog may mistakenly treat subjective users as dishonest ones,

Figure 5.9: Performance Comparison on the Epinions Dataset by Varying the Trust Threshold (F-Value)

Figure 5.10: Performance Comparison on the Epinions Dataset by Varying the Trust Threshold (Precision)

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 experience of the user.

Figures 5.9, 5.10, 5.11 and 5.12 present the performance of different approaches by varying the trust threshold. It shows that, in general, our model outperforms the other

Figure 5.11: Performance Comparison on the Flixster Dataset by Varying the Trust Threshold

Figure 5.12: Performance Comparison on the FilmTrust Dataset by Varying the Trust Threshold

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 to infer the (unobserved) (dis)trust links. This is especially important in online communities, where users are reluctant to explicitly identify their relationships with other users.

5.3 Summary

In this chapter, we propose a novel probabilistic graphical trust model, PGTM, separately considering *dishonesty* of advisors and their *subjectivity difference* with 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 compare our model with a baseline approach, and two state-of-the-art approaches including BLADE and Prob-Cog. Experimental results indicate that the latent variables in our model are both theoretically reasonable and computationally effective, and dishonesty and subjectivity difference are successfully distinguished. Further, we demonstrate 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.

CHAPTER 6

SubGroup: Learning From Users' Rating Behavior

Our PGTM model successfully deals with, and distinguish dishonesty and subjectivity difference issues when modeling advisors' trustworthiness. However, it ignores the fact that dishonesty and subjectivity overlap with each other to a certain extent. In the light of this, in this chapter, we present a novel SubGroup model for opinion evaluation.

Other than the dishonesty and subjectivity problems, most existing previous approaches (e.g. social network based methods and trust models) strive to model an advisor's trustworthiness for a user on the basis of either the common experience between the user and the advisor ([Teacy et al., 2006](#)) or the whole historic experience of them ([Zhang and Cohen, 2007](#)). Consequently, these models, on the basis of machine learning techniques, might fail since (1) the user and advisor have limited past experience in a community, especially limited common experience towards the same entities; and (2) a user's behavior is dynamic and evolving, which greatly increases the difficulty of tracking her behavior.

In view of the aforementioned problems, we propose a clustering-based method (SubGroup) that categorizes users into different groups with respect to their rating behavior. We distinguish subjective users from dishonest ones. In other words, users with similar subjectivity in rating entities are grouped together, and dishonest users are labeled as outliers. We further examine these dishonest users by dividing them into three different types. Consequently, on the one hand, each user cannot only directly employ other users' opinions in the same subjectivity group, but also effectively and rationally

use the opinions of those in different subjectivity groups (Etuk et al., 2013). On the other hand, the user might even take advantage of opinions provided by those dishonest advisors whose rating behavior follows definite patterns, while opinions of dishonest advisors with no static patterns would be ignored. Accordingly, users can maximally and effectively adopt advisors' opinions with our method. We prefer unsupervised learning (i.e. clustering) over supervised learning since we have no prior knowledge about the size and number of the subjectivity groups. Besides, the number of groups would also be varied as the dynamic change of users (new users join, and users might leave). The advantages of exploring group information through clustering instead of personal information (for opinion evaluation) can mitigate the gap between limited data of an individual user and sufficient data required for accurately learning the user's behavior. Moreover, we identify a set of features that can well capture the dynamic behavior of users.

Computationally, each user in the system is equipped with a software agent. Each agent clusters its user's advisors according to their historic experience. More specifically, a two-layered clustering approach is proposed for each agent to cluster advisors of its user into subjectivity clusters and three dishonest types (i.e. direct dishonest, disguise dishonest and misguidance dishonest). The first layer employs the DENCLU algorithm (Hinneburg and Keim, 1998) twice to identify advisors as either *subjective* or *dishonest*. Then, in the second layer, each advisor (not including misguidance dishonest ones) is assigned to two closest clusters with respective membership degrees¹². Given the clustering results, each agent further adopts a simple but effective group alignment algorithm that helps its user align advisors' ratings to the ones of her own. We conduct experiments on two different environments to validate the effectiveness of our approach: a distributed simulated e-marketplace for opinion evaluation, and three real datasets obtained from Epinions, Flixster and FilmTrust for rating prediction in recommender systems. Experimental results verify that our approach can help users better utilize ratings

¹²The sum of the membership degree for an agent in each group is 1.

provided by advisors.

The rest of the chapter is organized as follows. In Section 6.1, we summarize the definitions and studies of subjectivity and dishonesty, respectively, and then describe the SubGroup model in Section 6.2. In particular, first we introduce our research problem and procedural framework of the proposed method in Section 6.2.1, and then identify the features employed in the clustering algorithm in Section 6.2.2. Next, we describe the details of the proposed two-layered clustering approach in Section 6.2.3, and present a simple but effective group alignment algorithm in Section 6.2.4. In Section 6.3, we verify the effectiveness of our SubGroup model in comparison with trust models and recommender algorithms. Finally, we conclude this study in Section 6.4.

6.1 Subjectivity and Dishonesty

In this section, we summarize the definitions and studies of subjectivity and dishonesty, respectively.

6.1.1 Subjectivity

People are subjectively different. Opinions (or ratings) of each user in online communities imply a certain degree of subjectivity of this user. For example, in a rating system, user A rates a comic book as “5”, while user B rates the same book as “4”. In this case, we could conclude that A is a little more positive than B in this context of comic books. Hence, we choose to learn users' subjectivity from their rating behavior. Subjectivity analysis has been actively studied in various applications such as customers' opinion mining in online review forums and multi-document summarization ([Biyani et al., 2012](#)). Our research differs from those in the literature from two perspectives: (1) most previous research focuses on textual information, while we consider ratings; and (2) in the literature, subjectivity analysis is often defined as a binary-classification task, i.e. subjective or non-subjective. This is opposed to the clustering task in our research, where users are

clustered into multiple clusters (e.g. positive, neutral and negative in Noorian et al. (2011)), and users in the same subjectivity groups tend to provide similar ratings towards same entities. The number of subjectivity clusters is uncertain (≥ 2).

6.1.2 Dishonesty

Online environments open the door for dishonest users (i.e. attackers) to manipulate online rating systems by selfishly promoting or maliciously demoting certain entities (Feng et al., 2012a). According to the principle of veracity (Levine et al., 2010), users usually deceive for a reason, and that motives producing deception are usually the same as those that guide honesty. For example, in an e-commerce environment, ratings of an entity could not only reflect its popularity and reputation, but also greatly affect its sales. Hence, attackers tend to strategically manipulate their ratings in order to fulfill their goals (maximize their own profits or demote those of competitors).

Extending the work of Feng et al. (2012a), we focus on three types of dishonest users: (1) *direct dishonest* users consistently provide dishonest ratings to all entities. This is the most naive attacker model; (2) *indirect dishonest* users behave honestly (by providing honest ratings) to entities of certain types, but perform dishonestly to those of other types. They can also be called *disguise* users, since they disguise themselves as honest in some scenarios to gain trust of other users; and (3) *misguidance dishonest* users provide dishonest and honest ratings to entities following no static pattern. Dishonest users of this type are very difficult to track since their behavior is extremely sophisticated.

6.2 The SubGroup Approach

In this section, we present our approach in detail. Firstly, we introduce our research problem and procedural framework. Secondly, we identify the features employed in the clustering algorithm based on our intuitions and related work. Thirdly, we describe the details of the proposed two-layered clustering approach. Finally, we present a

simple but effective group alignment algorithm, which shows how the results of our clustering analysis can be effectively used in applications such as opinion evaluation and recommender systems.

6.2.1 Procedural Framework

In our approach, each user is equipped with a software agent that is responsible for managing the ratings of its own user and other users (advisors) towards entities in online communities. Each user has a set of past interactions with some entities and provides a rating in the range of $[0, 1]$ ¹³ for each interaction. We assume that a user u (equipping with an agent a) has previously interacted with a set of entities E . Based on these interactions (*rating behavior*), the user is described by a feature vector $F^m = \{f_1, f_2, \dots, f_m\}$, where m is the number of features. Similarly, agent a can also obtain the corresponding feature vectors for its user's advisors (*feature extraction*). The set of advisors is denoted by U^u . For implementation, U^u mainly involves the users who previously interacted with the same entities as user u . If u has limited historic interactions, agent a could expand U^u by also actively propagating requests for more information to the agents of other users until U^u is considerably extended. Then, agent a conducts our proposed *cluster analysis* towards u and advisors in U^u , which are clustered into different groups: either subjectivity groups or dishonest groups (outliers). Consequently, agent a can effectively utilize ratings provided by advisors given the results of the cluster analysis. The basic process is presented in Figure 6.1. As shown in the figure, we also describe an alignment algorithm that applies the outcome of the cluster analysis into various real applications, such as trust models for opinion evaluations, and recommender systems.

¹³Multi-nominal ratings in online communities could be normalized into the rating scale of $[0, 1]$.

Figure 6.1: Procedural Design of the Framework

6.2.2 Feature Identification

In this part, we aim to identify a feature vector $F^m = \{f_1, f_2, \dots, f_m\}$. Each feature for our research problem is expected to satisfy two objective requirements: (1) to maximize the distance between users of different subjectivity (including the dishonest ones); and (2) to minimize the distance between users of the same subjectivity (or dishonest type). To fulfil this goal, the most important issue is to understand the clues distinguishing the subjective behavior and dishonest behavior with regard to ratings (as presented in Section 6.1).

Following the reliability research in the medical domain (Sheline et al., 1999), we consider comparing ratings from different users with respect to two metrics: inter-user agreement and intra-user agreement. The former indicates the scenario where different users rate the same entity, while the latter refers to the scenario where a user provides ratings to different entities of the same quality. Honest users of the similar subjectivity have higher values on the two perspectives—provide similar ratings to the same entities as the group members, and consistently provide similar ratings to the products of same quality. Accordingly, subjective users of significantly different types might have lower

values on inter-user agreement, but higher values on intra-user agreement, while dishonest users have lower values on both perspectives. Additionally, another guideline for feature selection is: *rating behavior of subjective users is sort of consistent (might evolve with a low speed), while that of dishonest ones is comparatively unstable (strategically changing).*

With these guidelines in mind, in general, we first identify several propositions in the literature or by intuition, and then directly indicate some features related to these propositions for our research problem.

Proposition 1. *The rating difference between users in the same subjectivity group is much smaller than that between users coming from different groups (Lauw et al., 2008).* In other words, users in the same subjectivity group tend to provide similar ratings. Hence, the related features for user u are *variance of ratings*, *u 's average rating difference with other users regarding the same entities* (r_d^u), *variance of u 's rating difference with other users regarding the same entities* (v_d^u), etc.

$$r_d^u = \text{avg}_{i \in U^u, x \in E} (r_x^u - r_x^i) \quad (6.1)$$

$$v_d^u = \text{var}_{i \in U^u, x \in E} (r_x^u - r_x^i) \quad (6.2)$$

Proposition 2. *A user's rating of an entity may be subjectively impacted by other users' ratings (positive or negative) to the entity.* In this sense, the average or midpoint of ratings of the entity could be treated as a (quality) benchmark for the entity (Garcin et al., 2013). The associated features are *u 's average rating difference with the benchmark rating* (r_b^u), *variance of u 's rating difference with the benchmark rating* (v_b^u), etc.

$$r_b^u = \text{avg}_{x \in E} (r_x^u - \bar{r}_x) \quad (6.3)$$

$$v_b^u = \text{var}_{x \in E} (r_x^u - \bar{r}_x) \quad (6.4)$$

where $\bar{r}_x$ is the average rating towards entity x , i.e. benchmark rating.

Proposition 3. *The rating difference between two users would be relatively stable for two subjective users, but unstable for those where one or both of them are dishonest. Hence, corresponding features are the average of u 's rating difference with other users based on the commonly rated entities (r_c^u), the variance of u 's rating difference with other users based on the commonly rated entities (v_c^u), etc.*

$$r_c^u = \text{avg}_{i \in U^u, x \in E(u, i)} (r_x^u - r_x^i) \quad (6.5)$$

$$v_c^u = \text{var}_{i \in U^u, x \in E(u, i)} (r_x^u - r_x^i) \quad (6.6)$$

where $E(u, i)$ are the common entities between user u and i .¹⁴

Proposition 4. *For users having relatively few interactions, or lower interacting frequency with entities, they might be more reluctant to provide dishonest ratings. This is mainly because in this scenario, the dishonest ratings provided by them might have relatively less influence than those provided by other dishonest users (with a larger number of historic interactions or higher interacting frequency). One possible exception would be dishonest users having sybil identities. In this case, related features are the total number of interactions, skewness of ratings (r_s^u) (Doane and Seward, 2011), the interacting frequency, etc.*

$$r_s^u = \frac{\sum_{x \in E} (r_x^u - \bar{r}^u)^3}{(N - 1) \times s^3} \quad (6.7)$$

where $\bar{r}^u$, s and N are the mean, the standard deviation, and the number of the ratings provided by user u , respectively.

¹⁴Note that Proposition 1 considers each user's rating history, whereas Proposition 3 considers only partial of the history.

Proposition 5. *A user might vary her behavior under different contexts.* This proposition holds true for both dishonest users (e.g. the disguise dishonest users) and honest users (e.g. users might have different subjectivity in evaluating entities of different categories). Context information (e.g. category or time) is especially valuable in detecting dishonest behavior (Blair et al., 2010). Each user's features of this type might be *the variance of the user's rating for entities of different contexts (e.g. categories), the variance of the user's ratings in different time interval, etc.*

All the features are normalized to be in the range $[0, 1]$. Note that our objective of this section is not to figure out all the features or the best candidate features to achieve the best performance, but, more importantly, to *provide hints on identifying features for our research problem*. As the value of these features can mirror the change of users' behavior with the change of time, we argue that the clustering algorithm on the basis of these features is capable of well addressing users' dynamic and evolving behavior.

6.2.3 Cluster Analysis

We propose a two-layered clustering approach for each agent to cluster its user's advisors into different subjectivity groups and dishonesty types (outliers). The first layer mainly employs a density based cluster algorithm, DENCLU (Hinneburg and Keim, 1998), that crisply clusters users into different groups and outliers. The second layer is a fuzzy process which softly smooths and justifies the clustering results.

The First Layer: DENCLU

We choose DENCLU mainly because it can well deal with the environments with lots of noise, which allows us to flexibly address different scenarios, even those with a great proportion of dishonest users. Further, it has a firm mathematic basis (Hinneburg and Keim, 1998) instead of only heuristic clues. Moreover, the algorithm scales to relatively large sets of advisors, and hence can be effectively employed by each agent to build a

model of advisor grouping. This conforms with our research scenario where users can easily obtain a considerably large number of advisors as they interact with more entities.

DENCLU is based on the idea that the influence of each data point could be modeled formally using an influence function which describes the impact of the data point within its neighborhood. As mentioned in Section 6.2.1, for agent a , its user u and u 's neighborhood U^u (advisors) are described by the m -dimensional feature space F^m . For simplification, we denote the set of users including u and advisors in U^u as U^{u+} . The influence function is symmetric, continuous, and differentiable, such as the Square Wave function and Gaussian function (Hinneburg and Keim, 1998). In our case, we choose the Gaussian function since it should represent most real-world cases. The influence of user x regarding to u ($x \in U^{u+}$), $f_B^x(u)$, is defined as:

$$f_B^u(x) = e^{-\frac{d(u,x)^2}{2\sigma^2}} \quad (6.8)$$

where σ is a predefined threshold, and $d(u, x)$ is defined as the Euclidean distance¹⁵ between user u and x :

$$d(u, x) = \sqrt{\sum_{i=1}^m (f_i^u - f_i^x)^2} \quad (6.9)$$

Accordingly, the density function of user u is defined as the sum of the influence functions of u 's advisors. Thus, u 's density function $f_B^D(u)$ is defined as (Hinneburg and Keim, 1998):

$$f_B^D(u) = \sum_{x \in U^{u+}} f_B^x(u) = \sum_{x \in U^{u+}} e^{-\frac{d(u,x)^2}{2\sigma^2}} \quad (6.10)$$

On the basis of the density function, we then need to find all the density attractors x_j^* ($x_j^* \in U^{u+}$, where $j = 1, \dots, n_d$, and $n_d \leq \|U^{u+}\|$). Informally, density attractors are local maxima of the overall density function. If a point $x \in U^{u+}$ is *density-attracted* to a density attractor x_j^* , and x_j^* has a relatively bigger influence ($\geq \xi$, where ξ is a predefined bound), x belongs to cluster with x_j^* , otherwise, x is an outlier.

¹⁵The choice of distance functions is application-dependent. For example, for the application where users have many features with 0 values, the Hamming distance is more suitable.

The DENCLU algorithm consists mainly of two steps:

Step 1 (pre-clustering stage): the representation of U^{u+} is divided into m -dimensional hypercubes, with an edge-length of 2σ . Only hypercubes that contain at least one data point (i.e. one user) are determined, and called *populated cubes*. The set of populated cubes is denoted by C^u , where the number of hypercubes $\|C^u\|$ is sensitive to the choice of σ .

Step 2 (main stage – clustering stage): only the highly populated cubes (contain a pre-defined number of data points) and cubes which are connected to a highly populated cube, denoted as C_q , are considered in determining clusters. The basic process is the same as (Hinneburg and Keim, 1998).

The quality of DENCLU depends on a good choice of two parameters (ξ, σ) . σ impacts the influence of a user in her neighborhood, and ξ determines the minimum level for a significant density-attractor. Appropriate (ξ, σ) could help us to well deal with environments with different noise levels (i.e. different levels of the proportion of dishonesty users in the online community). Specifically, it is better to set ξ to be bigger than the noise level (defined in Hinneburg and Keim (1998)).

The process of our first layer is as follows: firstly, agent a conducts DENCLU towards U^{u+} with appropriate parameters (ξ_1, σ_1) , and then the output is subjectivity clusters $C_s = \{c_1^u, c_2^u, \dots, c_{u_s}^u\}$ where u_s is the number of subjective clusters and $u \in c_t$ ($t \leq u_s$), and dishonest users U_d^u (i.e. outliers) regarding to u .

Secondly, agent a further conducts DENCLU with new appropriate (ξ_2, σ_2) pair¹⁶ towards U_d^u by controlling the cluster number to be 2. Thus, two types of dishonest users i.e. *direct dishonest* users (c_{d1}^u) and *disguise dishonest* users (c_{d2}^u), are identified. $C_d^u = \{c_{d1}^u, c_{d2}^u\}$. The outliers of this round are considered as *misguidance dishonest* ones, and their ratings are **discarded** (or filtered out) for the next layer and applications.

¹⁶Here, the noise level is decided by the third type of dishonest users.

The Second Layer: Fuzzy Smoothing

For users around the border of each cluster, it might be over-positive to determine them as belonging to only one cluster. This is in accordance with the real-world scenario where some users have the mixed subjectivity of more than two distinct subjectivity groups. To resolve this problem, in the second layer, agent a conducts a *fuzzy smoothing process* towards users in U^{u+} (particular C_s^u and C_d^u).

For each user $u_c \in C_{c1}^u$ ($C_{c1}^u \in C_s^u, c1 \leq u_s$), we first compute the distance of u_c with each of other clusters in C_s^u (not including C_{c1}^u). The nearest cluster C_{c2}^u ($C_{c2}^u \in C_s^u$) to u_c is selected. Afterwards, u_c is considered as belonging to set C_{c1}^u and C_{c2}^u with membership $m_{c1}^{u_c}$ and $m_{c2}^{u_c}$ using *piecewise membership function*, respectively:

$$\begin{cases} m_{c1}^{u_c} = \frac{d(u_c, \text{mean}(C_{c2}^u))}{d(u_c, \text{mean}(C_{c1}^u)) + d(u_c, \text{mean}(C_{c2}^u))} \\ m_{c2}^{u_c} = 1 - m_{c1}^{u_c} \end{cases} \quad (6.11)$$

Similarly, for each user $u_c \in C_d^u$, she is considered as belonging to set C_{d1}^u and C_{d2}^u with membership $m_{d1}^{u_c}$ and $m_{d2}^{u_c}$ respectively:

$$\begin{cases} m_{d1}^{u_c} = \frac{d(u_c, \text{mean}(C_{d2}^u))}{d(u_c, \text{mean}(C_{d1}^u)) + d(u_c, \text{mean}(C_{d2}^u))} \\ m_{d2}^{u_c} = 1 - m_{d1}^{u_c} \end{cases} \quad (6.12)$$

Note that each agent conducts cluster analysis based on the historic interactions of its own user as well as those of other users in the user's neighborhood (advisors). For a user with limited experience, the agent could request the information (feature vectors) of advisors from other agents about which the agent already has good knowledge (evaluation).

6.2.4 Group Alignment

Agent a manages its own user u 's rating system, including the ratings provided by u 's advisors in C_s^u and C_d^u . The ratings provided by advisors of the misguidance dishonest type are directly discarded.

We propose a simple alignment algorithm for each agent to effectively adopt the ratings provided by advisors. We call it the *group alignment algorithm* mainly because we adapt the rating difference between two users according to the ratings of the other users in their clusters (see Equation 6.13). By doing so, each user could benefit from the collective knowledge of users in her group, and avoid the inaccuracy caused by limited information, which is the case for most users. Let us assume one user u belongs to two clusters c_{u1}^u and c_{u2}^u with respective membership degree m_{u1} and m_{u2} . And, there is another qualifying user $u_c \in C_s^u \cup C_d^u$ belonging to two clusters c_{c1}^u and c_{c2}^u with corresponding membership m_{c1} and m_{c2} . In this case, each rating r provided by u_c to entity x would be aligned to that of u , r_c^u , in the sense that adapts to u 's evaluation criterion:

$$r_u = r + \overbrace{\text{mean}(r_{c_{u1}^u}) * m_{u1} + \text{mean}(r_{c_{u2}^u}) * m_{u2}}^{u's \text{ group center}} - \overbrace{\text{mean}(r_{c_{c1}^u}) * m_{c1} + \text{mean}(r_{c_{c2}^u}) * m_{c2}}^{u'_c's \text{ group center}} \quad (6.13)$$

where $\text{mean}(r_{c_{u1}^u})$ is the average of ratings of users from c_{u1}^u to the entities of the same type as x . The right hand of Equation 6.13 (not including r) indicates the rating difference between u and u_c to an entity of the same type as x .

Each agent depends on the cluster analysis and group alignment approach to manage its own user's rating system. Our approach can be directly used in opinion evaluation in online communities (where in most MAS cases, trust models are employed), and recommender systems.

6.3 Experiments

We conduct experiments in two different environments, a distributed simulated e-marketplace for checking the effectiveness of our approach in opinion evaluation (trust models), and three real datasets obtained from Epinions, Flixster and FlimTrust for validating our approach for rating prediction in recommender systems.

6.3.1 Simulated E-marketplace

We examine the effectiveness of our approach in a distributed environment by conducting comparisons with the competing trust models.

Experimental Settings

We simulate an e-marketplace involving 55 sellers and 500 buyers. In our simulation, each seller is assigned a base reputation scaled from 0.5 to 1 with step of 0.05. For example, if base reputation of a seller is 0.5, the seller has a probability 0.5 of conducting transactions successfully. Buyers rate a seller with a value in the range $[0, 1]$. Buyers also have different subjectivity in rating their experience with same sellers. We define each buyer's subjectivity as a constant, which follows a Gaussian Distribution across all buyers and is in the range $[-1, 1]$. For example, if a buyer's subjectivity value is 0.5, it means that she would rate sellers as 1 if sellers' base reputation is bigger than 0.5. In each iteration, whether a buyer would conduct a transaction is decided by a predefined probability value. A buyer always seeks to conduct a transaction with the seller with the highest reputation value (computed using trust models) from her view.

In addition to our approach, we implement a baseline approach without considering subjectivity and dishonesty, but computes the reputation of sellers by directly averaging the ratings collected from other buyers for the sellers. We also choose to implement the TRAVOS model (Teacy et al., 2006), which is a good representative filtering approaches. As methods that consider both dishonest and subjective rating problem, the HABIT (Teacy

et al., 2012) method is chosen instead of PGTM (Fang et al., 2013a) and PRep (Haghpanah and desJardins, 2012) because the latter two are complicated to implement for our scenario (refer to the details in Chapter 2).

We evaluate their performance in computing the reputation of sellers. The performance of an approach is measured as the mean absolute error (MAE) between the reputation of sellers computed for each buyer using the approach, and the reputation of sellers using the ratings according to each buyer's own subjectivity (the ground truth about the reputation of sellers with respect to the buyer).

Results and Discussion

Here, we first check the effectiveness of each part of our approach. Then, we present the performance of our approach against our three benchmarks. We also examine these approaches in detail by varying the ratio of dishonest buyers in the simulated environment.

Model Effectiveness. In Figures 6.2 and 6.3, we analyze the impact of σ used in the first DENCLU algorithm (σ_1), and the second DENCLU algorithm (σ_2), respectively. Note that $\sigma_1 < \sigma_2$. Specifically, Figure 6.2 presents the performance of our method by varying σ_1 while fixing $\sigma_2 = 0.2$ and 0.3 , respectively. Figure 6.3 shows the performance of our method by changing σ_2 while fixing $\sigma_1 = 0.12$ and 0.17 , respectively. As can be seen, our method is sensitive to σ_1 , and relatively insensitive to σ_2 . With regard to different σ_2 , our method can consistently achieve the best performance when $\sigma_1 \simeq 0.17$. This implies that at that level of σ_1 , we can obtain the best number of the subjectivity groups (granularity level) for our rating alignment. Given that value of σ_1 , the MAE approximates to 0.18 , which demonstrates almost 50% improvement over the baseline method ($\text{MAE} \simeq 0.35$).

We also investigate the effectiveness of features used in our approach. We select 10 features to represent a buyer for our model according to propositions in Section 6.2.2, including *mean*, *the number*, and *standard deviation* of the buyer's ratings, *the rating differences with other users* and *the variance*, *rating differences with other users with*

Figure 6.2: Performance Change of Our SubGroup Model by Varying σ_1

Figure 6.3: Performance Change of Our SubGroup Model by Varying σ_2

regard to commonly rated sellers and the variance, rating differences with the benchmark ratings and the variance, and skewness of the buyer's ratings. Figure 6.4 demonstrates the performance of our method as the number of features changes. The x-axis represents the number of features used in the implementation. Given a specific number of features, the line shows the average performance of different feature combinations, and the error bars demonstrate the respective best and worst performance. As illustrated, the overall performance of our method increases as more features are considered, especially when the number of features reaches 7. This partly indicates that our propositions (see Section 6.2.2) are both reasonable and effective.

We further explore the effectiveness of our method by showing the performance of *our method without dishonesty clustering* (using DENCLU to cluster dishonest buyers in the

Figure 6.4: Performance Change of Our SubGroup Model by Varying Number of Features ($\sigma_1 = 0.12$, $\sigma_2 = 0.30$)

Figure 6.5: Performance Change of Our SubGroup Model with Different Component

first layer), and *our method without fuzzy smoothing*. The results are shown in Figure 6.5. As can be seen, we can conclude that both the fuzzy smoothing and the dishonesty clustering contribute to the performance of our method. This also represents that although some buyers (advisors) are dishonest, we can still extract valuable information from their opinions.

Model Comparison. Figures 6.6 and 6.7 show the performance comparisons between our approach and the other three approaches in both the basic and deceptive environments. For our approach, we set $\sigma_1 = 0.12$ and $\sigma_2 = 0.3$. From the results shown in Figure 6.6, we can see that our method performs consistently the best regardless of the number of interactions buyers have with sellers. HABIT performs better than TRAVOS, and both

Figure 6.6: Performance Comparison as the Change of Iterations

HABIT and TRAVOS perform much better than the baseline approach. Note that the marginal effect of new interactions on the performance of our method is very significant when there are only a few interactions for each buyer.

Based on the basic environment (number of iterations=100), we also examine the effect of deception as buyers lie about their past experience by conducting complementary lying behavior where if a true rating to a seller is r in the scale of $[0,1]$, the liar will modify the rating as $1 - r$. We vary the ratio of liars from 0 to 0.5, and plot the MAE results of different approaches in Figure 6.7. Our approach performs much better than the other approaches. Both our method and HABIT are not so much affected by lying buyers. This is mainly because in our simulations, the rating behavior of dishonest buyers is relatively static, and thus their ratings could still be used for our method. HABIT does, in part, also address this case because it also makes use of public information about advisors in inferring the properties of the dishonest buyers.

6.3.2 Real Environments

In this section, we validate the effectiveness of our approach in centralized real-world environments by comparing our approach with the state-of-the-art recommendation algorithms in rating prediction. We conduct these experiments using real datasets to compare our model with several competing models, including TidalTrust (Golbeck,

Figure 6.7: Performance Comparison When Varying the Ratio of Liars

2005), clustering-based recommender system (CoClustering) (George and Merugu, 2005), probabilistic matrix factorization (Mnih and Salakhutdinov, 2007), and SocialMF (Jamali and Ester, 2010)¹⁷.

Experimental Settings

Three real-world datasets are used in the experiments: Epinions, FilmTrust and Flixster. Users provide numerical ratings in the range of $[1, 5]$ on Epinions. Further, users can explicitly specify other users as trustworthy or not based on whether the ratings of others are consistently valuable or not for the user. We adopt the extended Epinions dataset where trust value is labeled as 1. We sample a subset by randomly selecting 5,000 users. The other two datasets are FilmTrust and Flixster where users can also indicate others as trustworthy, and provide item ratings ranging from 0.5 to 4.0 (5.0 in Flixster) with step 0.5. The statistics of the three datasets is given in Table 6.1.

The experiments are conducted by applying the leave-one-out technique; that is, each rating is iteratively hidden whose value will be predicted by applying our method, the TidalTrust, CoClustering, PMF, or SocialMF methods until all ratings in the datasets are tested. The performance is evaluated by two commonly used measures: the root

¹⁷We use the source codes of the latter three methods provided by MyMediaLite Recommender System Library (www.mymedialite.net), and adopt the corresponding settings suggested by the papers.

Table 6.1: The Statistics of the Three Datasets

Features	Epinions	Flixster	FilmTrust
users	5,000	5,000	1,508
items	376,458	13,527	2,071
trust	744	2,898	2,853
ratings	968,467	264,540	70,998
avg rating	4.6964	3.6560	3.0028

Table 6.2: Performance Comparison of Different Approaches

Methods	Epinions		Flixster		FilmTrust	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
SubGroup	0.4058	0.1046	0.9057	0.6824	0.8515	0.6566
TidalTrust	0.6759	0.5756	1.2449	0.9846	0.9687	0.7564
CoClustering	0.6398	0.3834	0.9263	0.6926	0.8472	0.6401
PMF	0.6498	0.4260	0.9264	0.6902	0.8384	0.6405
SocialMF	1.6427	1.3508	1.3749	1.0536	0.8602	0.6590

mean square error (RMSE) and mean absolute error (MAE)¹⁸. They both refer to the differences between the predictions and the ground truth, but differ from each other as indicated by their names. Generally, smaller RMSE and MAE values indicate better predictive accuracy.

Results and Discussion

Table 6.2 summarizes the performance comparisons between our approach and other approaches on three real datasets. As shown in Table 6.2, our approach achieves better performance than others in terms of both RMSE and MAE on the Epinions and Flixster

¹⁸MAE is a good indicator of average performance, while RMSE is more suitable when error distribution is Gaussian. It is often suggested that a combination of MAE and RMSE can better assess model performance (Chai and Draxler, 2014).

datasets, and a little worse but still comparably competitive to other approaches on the FilmTrust dataset. This validates the effectiveness of our approach with respect to rating prediction. The lack of good performance of our approach on FilmTrust dataset is possibly because: (1) each item in FilmTrust has a great deal of interactions with users (around 34), which is suitable for the implementation of other approaches; and (2) users are quite controversial (see Table 6.1). Hence, a considerable number of users are considered as *misguidance dishonest* and filtered out by our method. The performance of SocialMF and TidalTrust is much worse than PMF and CoClustering, which may be due to: (1) there exist noisy trust relationships in the three datasets; and (2) trust neighbors of a user might not share the same preferences of the user.

6.4 Summary

In this chapter, we propose a two-layered clustering approach, SubGroup, to address the advisors' subjectivity difference and dishonesty in providing opinions. Specifically, the agent of each user firstly clusters the advisors of its users into different groups, with respect to their rating behavior. And then, each advisor is assigned to two groups with respective membership degrees. Lastly, each agent adopts an alignment algorithm to help its user align advisors' ratings to the ones of her own. We conduct experiments on both a simulated distributed environment and real data (considered as centralized environments) collected from Epinions, Flixster and FilmTrust, respectively. Experimental results verify that: (1) our approach can better help users utilize ratings (opinions) provided by advisors, and is relatively robust to deceptive environments; (2) the identified features are validated to be effective for our research scenario, and each part of our approach contributes to performance improvement; and (3) our approach can achieve competitive performance in rating prediction with respect to recommender systems.

CHAPTER 7

Discussion

In this chapter, we discuss and compare our four models presented in Chapters 3-6, and briefly explore their potential to cope with other challenging scenarios such as resisting attacks. We then discuss how social factors can be applied in other scenarios, such as recommender systems.

7.1 Model Comparison

In this section, we firstly analyze and compare the effectiveness of the four models, and then briefly explore the potential of our models to resist various attacks.

7.1.1 Effectiveness Comparison

We perform a thorough comparison of our proposed four models: DiffTrust, SARC, PGTM, and SubGroup from the six perspectives of *information availability*, *subjectivity*, *dishonesty*, *dynamic*, *new users* and *accuracy*, where the last five are key factors for an effective trust model (Urbano et al., 2009, 2010). The basic results are summarized in Table 7.1.

- Information availability: as described in Chapter 4, information availability refers to the amount of available information required by different models for their successful implementation. In **DiffTrust**, each user computes her direct trust on others based on the shared interactions (*rating* information) between them, while the similarity between *contexts* (i.e. spatial and temporal information) can further

Table 7.1: Comparison of the Four Models

perspective	DiffTrust	SARC	PGTM	SubGroup
information availability	ratings+(context)	ratings+attributes	ratings+(trust)	ratings+(context)
subjectivity		✓	✓	✓
dishonesty	✓	—	✓	✓
dynamic	✓	—		✓
new users	✓		—	—
accuracy	✓	✓	✓	—

Note: ✓ means that the corresponding model can cope well with the target perspective, while — refers to that the model can *partially* address that perspective. Brackets around a type of information mean that the model still can work without that information, but will achieve better performance with it.

improve the accuracy of trust modeling. **SARC** learns two kinds of subjectivity of each user (advisor) based on a few interactions of the user with entities, where each interaction is presented by a *rating-review pair*. The corresponding review contains values of the objective attributes to describe that interaction. The experiments in Chapter 4 demonstrate that the requirement of detailed reviews and objective attributes is not very restrictive, and only a few rating-review pairs (e.g. 6 in the simulation) are needed for SARC to learn each user’s subjectivity relatively well. As described in Chapter 5, **PGTM** is modeled given information on *ratings* and *trust relationship* (if any), and experimental results indicate that it can achieve more accurate trust values than other state-of-the-art approaches without using the partially observable trust links. In common with DiffTrust, **SubGroup** can obtain satisfactory performance based on a few features extracted from users’ *rating* behavior.

- **Subjectivity:** refers to whether the models can deal with the subjectivity issue. As shown in Table 7.1, SARC, PGTM and SubGroup models emphasize the

subjectivity of users, while DiffTrust does not.

- Dishonesty: refers to whether these models can address the dishonesty issue. As shown in Table 7.1, DiffTrust, PGTM and SubGroup models address the dishonesty issue of advisors, while SARC does not. However, as validated by the experiments in Chapter 4, SARC is not dramatically affected by dishonest advisors as it treats them as the ones with different subjectivity.
- Dynamic: refers to whether these models are able to cope with the dynamic behavior of users (or advisors). **DiffTrust** can capture the dynamics of trust since the trustworthiness of an advisor from the view of a user is sensitive to the interactions between them, varied over time, influenced by other users and the environment, and also subject to the user's own propensity. **SubGroup** can also address the dynamic behavior of users, as the features for clustering analysis can mirror the change of users' behavior over time. **SARC** can be partially characterized by this perspective, as we can allow SARC to update each user's subjectivity based on her recent experience. On the contrary, **PGTM** might not act quickly to users' behavior change.
- New users: refers to whether these models can handle new users (advisors) who newly enter the online communities and have no past experience with entities. Of the four models, **DiffTrust** can cope with the trustworthiness of new advisors (or new users' trust on advisors) by considering the users' propensity to trust others and social proximity between them according to their profile information. **PGTM** and **SubGroup** can partially address the newcomer problem as PGTM can set a prior distribution for each parameter, while SubGroup can assume that new advisors come from the cluster with most advisors. In contrast, **SARC** cannot effectively handle the newcomer problem.
- Accuracy: refers to whether the trust models can accurately evaluate the trustworthiness of advisors for users. We acclaim that DiffTrust, SARC and PGTM

can accurately model an advisor's trustworthiness for each user, as presented in Chapters 3, 4, and 5. Although SubGroup could not very accurately measure an individual interaction between a user and an advisor, it can achieve more accurate evaluation for the whole community than other state-of-the-art approaches.

7.1.2 Robustness Comparison

Robust trust models are expected to be capable of resisting various attacks (Kerr and Cohen, 2009b,a; Jøsang and Golbeck, 2009; Zhang et al., 2012; Jøsang, 2012; Jiang et al., 2013). Here, we examine whether our trust models can defend against the following representative attacks, including *Constant* attack, *Camouflage* attack, *Whitewashing* attack, *Sybil* attack and *Collusion* attack. Constant attack is the simplest attack strategy (Zhang et al., 2012), where dishonest advisors constantly provide unfairly positive/negative ratings to entities. Camouflage attack refers to dishonest advisors acting strategically by camouflaging themselves as honest advisors at an early stage to build up their trustworthiness and then providing unfair ratings. In Whitewashing attacks, as new accounts can be freely and easily created, a dishonest advisor whitewashes its low trustworthiness by starting a new account with the initial trustworthiness value (higher than the previous low value), while in a Sybil attack, a *single* dishonest advisor creates multiple accounts to constantly provide unfair ratings to entities. On the contrary, Collusion attack refers to that *a set of* dishonest advisors cooperating together to provide untruthful ratings to mislead others (Jøsang, 2012). The overall analysis results are described in Table 7.2¹⁹.

Constant attack: as discussed in Table 7.1, DiffTrust, PGTM and SubGroup can handle the scenarios with dishonest advisors, which implies that they are robust to Constant attack. Accordingly, SARC might not be completely robust against the Constant attack, as it can only partially address the dishonest problem.

Camouflage attack: PGTM might be vulnerable to the Camouflage attack, since

¹⁹Note that, in our analysis, we mainly consider the malicious advisor behavior of providing untruthful (or unfair) ratings to entities and creating identities with fake profile information, and exclude other behaviors such as sharing untruthful intra-attribute subjectivity functions in SARC.

Table 7.2: Model Comparison in Resisting Attacks

attack type	DiffTrust	SARC	PGTM	SubGroup
Constant	✓	—	✓	✓
Camouflage	✓	—		✓
Whitewashing		✓	✓	✓
Sybil	—	✓	✓	✓
Collusion	✓	—	✓	

Note: ✓ means that the corresponding model is robust against the target attack, while — refers to that the model can only *partially* resist the attack.

those distribution functions are learned and updated using the whole past experience of each user (advisor), involving both honest and dishonest behavior. As SARC takes time to react to Camouflaging advisors' behavior change, it is not completely robust to the Camouflage attack before updating the changed subjectivity well. In contrast, DiffTrust and SubGroup can be robust against the Camouflage attack, as they can react more rapidly to the user (advisor) dynamic behavior.

Whitewashing attack: in the DiffTrust model, Whitewashing advisors can create fake identities, which causes users' social proximity with them to be incorrect. In this sense, DiffTrust might fail against the Whitewashing attack. The other three models are relatively robust to the Whitewashing attack.

Sybil attack: in the four models, we consistently use each user's direct experience (local knowledge) with advisors to evaluate their trustworthiness (or advisor). In this way, they are robust against the Sybil attack. However, in the DiffTrust model, the profile information of some identities created by Sybil attackers might be fake, which would decrease the performance of the model, especially in the early stage of those identities' life cycle in the online communities.

Collusion attack: SARC can partially address the collusion attack as it learns each user's subjectivity according to her own past experience. Besides, it treats all those colluding dishonest advisors as having subjectivity difference. We claim that

both DiffTrust and PGTM can address the Conclusion attack problem, as both of them can accurately model the trustworthiness of advisors. In this case, even if the majority of advisors provide untruthful ratings to an entity, we can still have a relatively wise measurement on the target entity's reputation (or quality). Conversely, SubGroup might fail to defend against the Collusion attack as each user might not be able to accurately evaluate her own trust on colluding advisors, as described previously.

In the future, to validate our analysis results, we will carry out detailed experiments to demonstrate the robustness of the four models against these five representative attacks. More details can be found in Chapter 8.2.3.

7.2 Adapting Trust Factors into Recommender Systems

In this section, we explore the possibility of applying trust-related social factors into other areas. In particular, we use recommender systems as an example, and demonstrate that trust-related social factors can be employed in current trust-aware recommender systems to further improve their performance in terms of recommendation accuracy.

Trust has been extensively exploited for improving the predictive accuracy of recommendations by ameliorating the issues such as *data sparsity* and *cold start* that recommender systems inherently suffer from (Adomavicius and Tuzhilin, 2005; Massa and Avesani, 2007; Ma et al., 2009; Chowdhury et al., 2009; Ray and Mahanti, 2010; Jamali and Ester, 2010; Guo et al., 2012). In essence, trust provides additional information from which user preference can be better modelled, complementary to rating-based similarity. Both implicit (O'Donovan and Smyth, 2005) and explicit (Massa and Avesani, 2007; Chowdhury et al., 2009; Ma et al., 2009; Ray and Mahanti, 2010; Jamali and Ester, 2010; Guo et al., 2012) trust have been investigated in the literature. The former is usually inferred from user-item interactions (i.e. ratings) whereas the latter is directly specified by users indicating whom and to what extent they trust. In contrast, although distrust is recognized to play an equivalently important role as trust (McKnight and

Choudhury, 2006), the investigation of utilizing distrust in recommender systems is still in its infancy (Victor et al., 2011, 2013). To the best of our knowledge, no prior work has attempted to predict distrust for improving recommender systems.

Another issue of existent trust-aware recommender systems is the simplified modelling of trust as a concept with a single aspect, such as the ability to provide accurate ratings (known as *competence*) (O'Donovan and Smyth, 2005) or the probability of behaving maliciously. However, it is well acknowledged in Social Science that trust is a concept with multi-faceted properties (Mayer et al., 1995; McKnight and Chervany, 2001). One possible explanation is that only limited information is available in the few and publicly accessible datasets. Although some efforts have been made to capture multiple aspects (e.g. information credibility (Kwon et al., 2009)) of raters (who give ratings) in recommender systems, they are essentially distinct concepts from trust. A generally agreed proposition states that people trusting each other may not always share similar preferences (Jøsang et al., 2011). This statement leads to the following interesting research questions: (a) *which aspects of trust reflect user preferences more and hence should be more considered for user preference modelling?* and (b) *how to make an effective decision on whether to incorporate a user for recommendation in terms of trust scores?* Answers would provide guidance on whom and to what extent one can trust, especially given the fact that most available (i.e. explicit) trust scores are binary, i.e., either 1 (trust) or -1 (distrust) without specific degrees of trust or distrust.

In this section, we aim to address these research questions by proposing a framework of trust and distrust, taking into consideration both interpersonal and impersonal aspects of trust and distrust adapted from Social Science (McKnight and Chervany, 2001). Specifically, four interpersonal aspects (i.e. benevolence, competence, integrity and predictability)²⁰ are computationally modelled based on users' past ratings, while impersonal aspects (e.g. degree centrality) are formulated from the perspective of social links in trust networks. Note that the social links in a trust network consist of both trust

²⁰Three of them have also been used in the proposed PGTM model in Chapter 5.

and distrust connections among users. Two logistic regression models are developed and trained by accommodating these factors and then applied to predict continuous values of users' trust and distrust, respectively. Of the four interpersonal aspects, two (benevolence and competence) are shown to be positively correlated with trust, whereas two (integrity and predictability) are negatively correlated. In addition, competence and predictability are positively correlated with distrust whereas integrity and benevolence are negatively correlated, confirming the assumption that trust and distrust are related but distinct constructs. We further refine the trust information using the predicted distrust information. These newly generated trust values can then be applied in trust-aware recommender systems. The experimental results on real-world datasets demonstrate the effectiveness of our proposed model for improving the performance of three representative trust-aware recommendation algorithms. In addition, the generality of our model is also empirically demonstrated. In all, our work is the first to comprehensively study the multiple aspects of trust and distrust in the context of recommender systems. The results lead to refined trust and distrust predictions, and offer notable improvements in recommendation accuracy when these predictions are utilized.

The rest of the section is organized as follows. Section 7.2.1 elaborates the proposed (dis)trust framework, and Section 7.2.2 introduces the trust and distrust prediction models. The effectiveness of our approach is evaluated and discussed in Sections 7.2.3 and 7.2.4, respectively.

7.2.1 The (Dis)Trust Framework

In this section, we introduce the formal definitions of the interpersonal and impersonal aspects of trust and distrust from which they will be computationally modeled according to users' historic ratings and trust networks.

Both trust and distrust are well-known as heterogeneous rather than homogenous concepts in the fields of social science and computational trust, each of which is composed of multiple aspects [Mayer et al. \(1995\)](#); [McKnight and Chervany \(2001\)](#). Trust

Figure 7.1: The Proposed (Dis)trust Framework

consists of three major parts, namely dispositional trust, institution/structural-based trust, and interpersonal trust (McKnight and Chervany, 2001). Dispositional trust, also known as a trustor's trust propensity, refers to the trustor's inherent propensity to trust other users. Mathematically, it could be treated as a constant (in the range of $[0, 1]$) subject to each trustor. Institution/structural-based trust refers to a belief held by a trustor about impersonal things of a trustee such as environments and situation. Hence, in our framework, as all users are in the same environments, we differentiate this part of the trustee by regarding it as trustor's public view of the trustee's trustworthiness. This is mainly determined by impersonal aspects of the trustee such as her reputation and position in a trust network. The impersonal aspects also have an impact on trustor's perception and hence the trust evaluation (McKnight and Chervany, 2001). Interpersonal trust mainly involves *benevolence*, *integrity*, *competence*, and *predictability*.

With respect to the original trust model in (Mayer et al., 1995; McKnight and Chervany, 2001), we make a minor modification towards the connections between the aspects and trust as shown in Figure 7.1. Specifically, we regard the combination of each aspect of a trustee and the propensity of a trustor as *an aspect of the trustee perceived by the trustor*, or a *trusting belief* of the trustor that the trustee has the corresponding characteristic in her favor. Therefore, trust in our model is connected with four different trust beliefs (interpersonal aspects), each of which is regarded as a trust aspect of a trustee perceived

by a trustor. Together with the trustor's trust propensity and impersonal aspects of the trustee, these aspects are known as the antecedents of trust (Mayer et al., 1995; McKnight and Chervany, 2001), and elaborated as follows.

- **Benevolence** refers to the extent to which a trustee cares about the preferences of a trustor (McKnight and Chervany, 2001), i.e., the willingness of the trustee to do good deed for the trustor. For the user with whom the trustor has a high benevolence belief, her preferences are more likely to be similar to those of the trustor. In our case, it means that both users report similar ratings on many items.
- **Integrity** refers to the extent to which a trustee conforms to a norm or code of moral or artistic values (Mayer et al., 1995). It stresses the characteristic of the trustee to follow the norm or rules of an organization, and to have a core set of values to guide behavior. To put it simply, the trustor believes that the trustee will always keep good-faith agreements, tell the truth, act ethically and fulfill its promises (McKnight and Chervany, 2001). In contrast to benevolence, integrity is more concerned with the characteristic of the trustee than the trust relationship (McKnight and Chervany, 2001).

The aspects of benevolence and integrity are somewhat complementary to each other in evaluating the trustworthiness of a specific trustee. Specifically, although benevolence shows the honesty or willingness of a trustee towards a trustor, it may fail to work in some scenarios where only limited interactions between the two users exist. This issue can be partially addressed by the integrity via considering the experience of all the users. Similarly, in the cases where integrity tends to be misleading, e.g., when most users are malicious, benevolence can help cope with this issue by relying more on personal experience between the two users.

- **Competence** refers to the ability or the power of a trustee to conduct the actions that are expected by a trustor in a specific domain (Mayer et al., 1995). Hence, competence is domain (context)-specific. For example, a user providing satisfying

recommendations of purchasing cars may not be an expert in buying clothes. In other words, the user receiving a high competence belief from the trustor is capable of providing satisfactory recommendations to the trustor in a specific context. The more experience the trustee has in the specific context, the more competent she will be in the view of the trustor.

- **Predictability** refers to the consistency of a trustee's actions (good or bad, negative or positive) such that the trustor can make a prediction in a given situation ([McKnight and Chervany, 2001](#)). Different from integrity, the value of predictability is neutral. Specifically, users' high predictability could mean that they always provide relatively high or low recommendations needed by the trustor, or consistently meet the trustor's preferences. Predictability is able to alleviate the problem of behaviors changing strategically; that is, a user may first act honestly but conduct dishonest behaviors later.
- **Impersonal aspects** represent different situations a trustor may encounter when interacting with a trustee. In our framework, they summarize the aspects of a trustee from the public view, which are independent of the interpersonal relationship between trustor and trustee. The representative information includes trustee's reputation, position in the trustor network, degree centrality ([Opsahl et al., 2010](#)), authority, and even their profile information.

As mentioned above, distrust is recognized as a distinct construct and opposed to trust. Trust and distrust may exist simultaneously between a trustor and a trustee. Distrust is also a multi-faceted concept, and is formalized as the mirror image of the trust concept ([McKnight and Chervany, 2001](#)). Similarly, we connect the distrust with the aspects identified in the framework, which is illustrated in Figure 7.1.

7.2.2 Trust and Distrust Prediction

In this section, we firstly formulate the (dis)trust aspects based on users' historical experience. Then, we present two logistical regression models by accommodating these aspects to predict continuous values of users's trust and distrust, respectively. Finally, we further refine the trust links given the predicted trust and distrust values.

Formulations of Aspects

Given the formal definitions, we proceed to formulate the four aspects in the light of users' historical experience (i.e. ratings). For clarity, we first introduce a number of notations. Suppose there are two users: a trustor a and a trustee b , and each user has a set of experiences denoted by E_a and E_b , respectively. An experience is denoted by a 5-tuple $e_u = (u, j, r_{u,j}, t, c)$, indicating that a user u rated item j with a rating $r_{u,j}$ at time t under context c . Hence, users' experiences can be represented as $E_a = \{e_{a1}, \dots, e_{am}\}$ and $E_b = \{e_{b1}, \dots, e_{bn}\}$, where m and n are the number of experiences of users a and b , respectively.

Based on user experience, we then model the four general trust aspects (i.e. beliefs, see Figure 7.1) of trustee b from the viewpoint of trustor a , as well as the trust value that trustor a has towards trustee b . Note that belief could be modeled by evidence ([Shafer, 1976](#)). Following the definitions described in Section 7.2.1, we model the four aspects as follows.

- Benevolence, $Be(a, b)$. As benevolence refers to the closeness of shared experiences between two users a and b , it is modeled as the user similarity which is usually used in collaborative filtering and computed by the Pearson correlation coefficient ([Adomavicius and Tuzhilin, 2005](#)):

$$Be(a, b) = \frac{\sum_{j \in E_{a,b}} (r_{a,j} - \bar{r}_a)(r_{b,j} - \bar{r}_b)}{\sqrt{\sum_{j \in E_{a,b}} (r_{a,j} - \bar{r}_a)^2} \sqrt{\sum_{j \in E_{a,b}} (r_{b,j} - \bar{r}_b)^2}}, \quad (7.1)$$

where $E_{a,b} = E_a \cap E_b$ is the set of shared experience on the commonly rated items between users a and b , and $\bar{r}_a, \bar{r}_b$ are the average of the ratings reported by users a and b , respectively. Alternative similarity measures such as cosine similarity (Adomavicius and Tuzhilin, 2005) could also be applied.²¹

- Integrity, $In(b)$. Integrity is independent of the trustor-trustee relationship, and hence it is formulated based on the past experience of the trustee regardless of the trustor's actions and evaluation. Specifically, the behaviors of the majority are treated as the norm or the code when evaluating the integrity of the trustee; i.e., the similarity between the trustee's behaviors and the majority's. Hence, integrity is computed by the similarity between the preferences of the trustee and the average:

$$In(b) = \frac{\sum_{j \in E_b} (r_{b,j} - \bar{r}_b)(\bar{r}_j - \bar{r})}{\sqrt{\sum_{j \in E_b} (r_{b,j} - \bar{r}_b)^2} \sqrt{\sum_{j \in E_b} (\bar{r}_j - \bar{r})^2}}, \quad (7.2)$$

where $\bar{r}_j$ refers to the average of the ratings on item $j \in E_b$, and $\bar{r}$ is the average of the ratings on all items.

- Competence, $Co(a, b, c)$. The competence of the trustee b is described from the viewpoint of the trustor a under a specific context c . Two factors are taken into account: the number of user b 's experience under context c (see Equation 7.4), and the ratio of correct recommendations given by user b to all the other users in the system (see Equation 7.3), employing the basic idea of O'Donovan and Smyth (2005). Competence is computed by integrating both factors:

$$Co(a, b, c) = \gamma \frac{\sum_{j \in E_b} \sum_{u \in U_j} count(|r_{b,j} - r_{u,j}| < \varepsilon)}{\sum_{j \in E_b} \|U_j\|}, \quad (7.3)$$

where U_j represents the set of users who have a piece of experience about item j , and ε is a predefined error tolerance threshold below which a rating $r_{b,j}$ of the

²¹Note that this also holds for the computation of the integrity (see Equation 7.2).

trustee b is treated as a correct recommendation for item j relative to the other's real preference $r_{u,j}$. And γ is defined by:

$$\gamma = \begin{cases} \frac{N_{b,c}}{N_c^a} & \text{if } N_{b,c} \leq N_c^a; \\ 1 & \text{otherwise;} \end{cases} \quad (7.4)$$

where $N_{b,c}$ is the number of experience under context c out of the total m experience that user b has, and N_c^a is the minimal number of experience under context c required by the trustor a such that a user can be regarded as a reliable recommender.

- Predictability, $Pr(a, b)$. Different from integrity, the predictability of trustee b is defined as the degree to which the (positive, neutral or negative) trend of b 's rating behaviors is distinct from that of trustor a . Formally, it is computed by:

$$\begin{aligned} n_u &= \text{count}_{j \in E_{a,b}}(|r_{a,j} - r_{b,j}| \leq \theta); \\ n_n &= \text{count}_{j \in E_{a,b}}(r_{a,j} - r_{b,j} > \theta); \\ n_p &= \text{count}_{j \in E_{a,b}}(r_{a,j} - r_{b,j} < -\theta); \\ Pr(a, b) &= \frac{\max(n_u, n_p, n_n) - \min(n_u, n_p, n_n)}{\|E_{a,b}\|}, \end{aligned} \quad (7.5)$$

where n_u , n_n and n_p refer to the neutral, negative and positive trends of user b 's rating behaviors compared to trustor a 's behaviors, respectively; θ is a threshold predefined by trustor a . The intuition is that for a user who is highly predictable, the difference in trends should be significant. In case of $n_u = n_n = n_p$, we obtain the lowest predictability since it is difficult to predict the next behavior of the trustee.

- Impersonal aspects: we identify four kinds of impersonal aspects in our computational model on the basis of the degree of a trustee in the trust network. The degree, as one of the centrality measurements, essentially records the aggregate public relations of the trustee in the network. The four aspects based on degree of trustee b are trust indegree $d_{in}^+(b)$, trust outdegree $d_{out}^+(b)$, distrust indegree $d_{in}^-(b)$

and distrust outdegree $d_{out}^-(b)$, referring to trustee b 's *incoming trust links*, *outgoing trust links*, *incoming distrust links* and *outgoing distrust links* respectively.

Trust Prediction

For trust prediction, we define $t_{a,b,c} \in [0, 1]$ as the trust value that trustor a has towards trustee b under context c , where 0 means a complete lack of trust and 1 completely trusted. The trust value will be influenced by the set of eight aspects that we investigated, denoted by $A(a, b) = \{Be(a, b), Co(a, b, c), In(b), Pr(a, b), d_{in}^+(b), d_{out}^+(b), d_{in}^-(b), d_{out}^-(b)\}$. In practice, users may specify other users as trusted neighbors ($t = 1$), whereas if trustor a has no direct trust link to trustee b , we consider that a has no trust towards b ($t = 0$). The trust and absence of trust connections will help build a useful model of the trust aspects and the overall trust. Specifically, the expected probability²² that trustor a completely trusts the trustee b under context c (denoted as $p^+(a, b, c)$) can be written as:

$$p^+(a, b, c) = E(t_{a,b,c} = 1 | A(a, b)). \quad (7.6)$$

We apply logistic regression to classify trust from lack of trust, and obtain the importance weight of each aspect related to trust. To be specific, the logit of the probability is modeled as a linear combination of $A(a, b)$ (Liu et al., 2012):

$$\text{logit}(p^+(a, b, c)) = \log\left(\frac{p^+(a, b, c)}{1 - p^+(a, b, c)}\right) = \alpha_0^{a+} + (\alpha_A^{a+})^T \cdot A(a, b), \quad (7.7)$$

where $\alpha_A^{a+} = \{\alpha_1^{a+}, \alpha_2^{a+}, \alpha_3^{a+}, \alpha_4^{a+}, \alpha_5^{a+}, \alpha_6^{a+}, \alpha_7^{a+}, \alpha_8^{a+}\}$, and α_0^{a+} is interpreted as the intrinsic trust propensity of the trustor a . Then the probability $p^+(a, b, c)$ is derived by:

$$p^+(a, b, c) = \frac{1}{1 + e^{-(\alpha_0^{a+} + (\alpha_A^{a+})^T \cdot A(a, b))}}. \quad (7.8)$$

Based on the trust information directly specified by real users, we are able to train

²²The probability is thus treated as the trust value.

this model and learn the coefficients; i.e., the importance weight of each aspect related to trust. The weights α_A^{a+} can be used to compute implicit, or refine explicit trust values from user experience.

Distrust Prediction

Following the process of trust prediction, the expected probability that the trustor a completely distrusts the trustee b under context c (denoted as $p^-(a, b, c)$) can be written as:

$$p^-(a, b, c) = E(d_{a,b,c} = 1 | A(a, b)), \quad (7.9)$$

where $d_{a,b,c} = 1$ represents that a completely distrusts b under context c . We also apply logistic regression to classify distrust from a lack of distrust, and obtain the importance weight of each aspect related to distrust:

$$\text{logit}(p^-(a, b, c)) = \log\left(\frac{p^-(a, b, c)}{1 - p^-(a, b, c)}\right) = \alpha_0^{a-} + (\alpha_A^{a-})^T \cdot A(a, b), \quad (7.10)$$

where $\alpha_A^{a-} = \{\alpha_1^{a-}, \alpha_2^{a-}, \alpha_3^{a-}, \alpha_4^{a-}, \alpha_5^{a-}, \alpha_6^{a-}, \alpha_7^{a-}, \alpha_8^{a-}\}$, and α_0^{a-} is interpreted as the intrinsic distrust propensity of the trustor a . Then the probability $p^-(a, b, c)$ is derived by:

$$p^-(a, b, c) = \frac{1}{1 + e^{-(\alpha_0^{a-} + (\alpha_A^{a-})^T \cdot A(a, b))}}. \quad (7.11)$$

Based on the distrust information directly specified by real users, we are able to train this model and learn the importance weight of each aspect related to distrust. The weights can be used to compute implicit, or refine explicit distrust values from user experience.

Trust Link Refinement

Given the predicted probability of complete trust $p^+(a, b, c)$ and distrust $p^-(a, b, c)$ according to Equations 7.8 and 7.11, we can further refine the trust links by filtering out the

possibly inaccurate trust or distrust links using the following rules:

$$\begin{aligned}
 &\text{if } p^+(a, b, c) > p^-(a, b, c), \text{ trust link from } a \text{ to } b; \\
 &\text{if } p^+(a, b, c) < p^-(a, b, c), \text{ distrust link from } a \text{ to } b; \\
 &\text{if } p^+(a, b, c) = p^-(a, b, c), \text{ no link from } a \text{ to } b.
 \end{aligned} \tag{7.12}$$

Furthermore, we could also refine the trust degree using Equation 7.13 for other specific purposes such as comparing the trust degrees between different user pairs.

$$t(a, b, c) = \begin{cases} p^+(a, b, c) - p^-(a, b, c) & \text{if } p^+(a, b, c) > p^-(a, b, c) \\ 0 & \text{otherwise.} \end{cases} \tag{7.13}$$

7.2.3 Evaluation

For evaluation, we aim to explore the effectiveness of our proposed (dis)trust framework by incorporating the generated trust information into three representative trust-aware recommender systems.

Datasets

Three real-world datasets (Epinions, FilmTrust and Flixster) are used in the experiments. Epinions enables users to review products by adding text comments and issuing numerical ratings in the range of $[1, 5]$. Further, users can explicitly specify other users that they trust (to the trust list) or distrust (to the block list) based on whether the reviews and ratings of others are consistently valuable or not for the user. We adopt the extended Epinions dataset where trust value is labeled as 1 and distrust as -1 . We sample two subsets by randomly selecting 5,000 and 10,000 users, respectively. The other two datasets are FilmTrust and Flixster where only trust exists and no distrust information is available. Users can only indicate others as trusted, and provide item ratings scaled from 0.5 to 4.0 (5.0 in Flixster) with step 0.5. The statistics of the four datasets is presented in Table 7.3.

Table 7.3: The Statistics of the Four Datasets

Features	Epinions1	Epinions2	FilmTrust	Flixster
users	5,000	10,000	1,508	5,000
items	376,458	519,491	2,071	13,527
trust	744	3,443	2,853	2,898
distrust	424	1,398	n.a.	n.a.
ratings	968,467	2,017,158	70,998	264,540
avg rating	4.6964	4.6863	3.0028	3.6560

Experimental Settings

Since the two Epinions subsets are the only available collections that contain both trust and distrust information, we use them to train two logistic regression models for trust and distrust respectively. Specifically, the users who specify both trust and distrust statements to others are selected as the training data in order to learn the coefficients (i.e. the importance weights) of each trust and distrust aspect according to Equation 7.7 and Equation 7.10, respectively. Due to the limitations of available datasets, we do not take into account the context information in the experiments. Further, we set $\varepsilon = 0.1$ for competence (see Equation 7.3) and $\theta = 0.1$ for predictability (see Equation 7.5) computations. Although the other datasets (FilmTrust and Flixster) do not contain distrust information (and hence cannot train a regression model independently), they may be useful in testing the effectiveness of these aspects by adopting the models learned from the Epinions datasets. The intuition is that although the exact or absolute coefficient values may vary across datasets, the relative importance weights may follow the same trends for key factors. To be specific, we apply the coefficients learned from Epinions1 to FilmTrust, and those learned from Epinions2 to Flixster according to the comparative sizes of the corresponding datasets.

After obtaining the aspect coefficients, we regenerate or predict the trust values in the light of different combinations of the two types of trust and distrust aspects, and in

total we obtain 3 different combinations and their corresponding trust values. Hence, the effectiveness of the new trust information (refined by the predicted distrust information) can be investigated by the recommendation performance in comparison with the original ones. Specifically, to demonstrate the effectiveness, we adopt three representative trust-aware algorithms to generate recommendations:

- **TidalTrust**, proposed by [Golbeck \(2006\)](#), uses trust values to substitute user similarity to weigh user ratings when generating recommendations.
- **Merge**, proposed by [Guo et al. \(2012\)](#), incorporates the ratings of trusted neighbors to form a more complete rating profile for active users, where the trust propagation length is 1.
- **SocialMF**, proposed by [Jamali and Ester \(2010\)](#), considers the trust information and propagation of trust information into the matrix factorization model for recommender systems. In our experiments, we adopt the same settings of parameters as suggested in ([Jamali and Ester, 2010](#)), and source code provided by MyMediaLite recommender system library²³.

To have a better understanding of the effectiveness, we split each dataset into three different views in terms of item-related properties as used in ([Guo et al., 2012](#); [Massa and Avesani, 2007](#)):

- **All** represents the whole dataset.
- **Controversial Items** are those items which received ratings with standard deviation greater than 1.5.
- **Niche Items** are those items which received fewer than 5 ratings.

The experiments are conducted by applying the *leave-one-out* technique, that is, each rating is iteratively hidden whose value will be predicted by applying the TidalTrust,

²³<http://www.mymedialite.net>

Table 7.4: The Coefficients of Trust Aspects

Aspect \ Dataset	Epinions1		Epinions2	
	trust	distrust	trust	distrust
benevolence	0.772	-1.2295	0.6332	-1.4537
competence	2.3706	0.988	2.5458	1.6816
integrity	-0.5816	-0.1122	-0.6597	-0.461
predictability	-0.0471	0.313	-0.3666	0.5724
trust indegree	-0.055	0.0159	-0.0387	-0.0006
trust outdegree	0.0615	-0.0042	0.0677	0.011
distrust indegree	-0.0765	-0.3697	0.0016	-0.2533
distrust outdegree	-0.0125	0.2347	-0.0066	0.144

Merge, or SocialMF method until all ratings in the datasets are tested. The performance is evaluated by two commonly used measures: the root mean square errors (RMSE) and mean absolute errors (MAE). They both refer to the differences between the predictions and the ground truth, but differ from each other as indicated by their names²⁴. Generally, smaller RMSE and MAE values indicate better predictive accuracy.

7.2.4 Results and Analysis

The experimental results are presented in two-fold: (1) the importance weights of the trust and distrust aspects learned from logistic regression models; and (2) the effectiveness of the trust and distrust aspects applied in recommender systems in comparison with that of the original trust values.

Importance of Trust and Distrust Aspects

We use the L2-regularized logistic regression provided by LIBLINEAR²⁵ to train the data of Epinions1 and Epinions2. The coefficients (i.e. the importance weights) of the trust and distrust aspects are illustrated in Table 7.4. Note that since the implementation of LIBLINEAR tends to minimize the bias part (to 0) during the model fitting process, we do not present the results of the aspect about trustor's propensity. In fact, its value is often equal to or very close to 0. Further, since logistic regression has a strict requirement on the sample size, we adopt a well-known rule of thumb, i.e., the **1 in 10 rule** (Harrell et al., 1984) to specify a minimum size of the sample for reliable training. In particular, a minimum number 10 of trust (or distrust) links are required for each aspect, that is, at least 80 trust (or distrust) examples are required in order to obtain a reliable model. However, we find that only few users in the training sets could meet this requirement. Therefore, we train the logistic regression models based on the trust and distrust networks of all the users and adopt the learned coefficients for all the users. An alternative method is to divide users into different clusters according to user similarity and then the coefficients could be learned using all the users' experience within the same cluster. This may lead to more accurate coefficients for similar users. Although we do not conduct our experiments in this way in the current work, we demonstrate that our method based on the logistic regression models learned from all users (i.e. general knowledge) could already significantly improve the recommendation accuracy.

Table 7.4 shows that consistent results for the four interpersonal aspects²⁶ with trust and distrust are obtained in both Epinions1 and Epinions2 datasets. In general, benevolence and competence are both positively correlated with trust whereas integrity and predictability are negatively correlated. In other words, the first two aspects are

²⁴MAE is a good indicator of average performance, while RMSE is more suitable when error distribution is Gaussian. It is often suggested that a combination of MAE and RMSE can better assess model performance (Chai and Draxler, 2014).

²⁵<http://www.csie.ntu.edu.tw/~cjlin/liblinear/>

²⁶Their values are in the range [0,1], while the values for the four impersonal factors are integers (≥ 0).

Figure 7.2: The Comparison of Refined Trust Links with the Original Ones

more likely to increase the probability of trust, but the latter two decrease the probability. More specifically, competence shows the greatest correlation with trust, followed by benevolence. This may imply that users in recommender systems are more concerned with personal experience (e.g. benevolence) rather than collective opinions (e.g. integrity) when establishing trust. Further, a person whose behaviors are highly predictable does not guarantee high trustworthiness in trust building because predictability is value-neutral. In contrast, competence and predictability present positive correlation with distrust whereas benevolence and integrity are negatively correlated with distrust. It should also be noted that the result for each individual impersonal aspect is not very consistent across two datasets. This might be due to the fact that we only capture partial trust and distrust information for users in our datasets, as we only consider the trust and distrust information of each user to our sampled users. Overall, however, the coefficients for impersonal aspects could still be considered as consistent in the sense that the aggregated effect of the trust network related impersonal aspects (trust indegree and outdegree) is positive, while that of the distrust network related impersonal aspects is negative, for both trust and distrust. This could be partially explained because a trustee with more trusted and trusting neighbors could have more far-reaching influence on other users, and thus are either more trusted or distrusted by others. In other words, a trustworthy user would be considered as more trustworthy by trustors, and further trusted by more people, and vice versa.

Table 7.5: Performance Comparison Based on Refined Trust Using Epinions1

Methods	Aspects	All		Controversial Items 2,777 users 7,242 ratings		Niche Items 4,705 users 539,881 ratings		FilmTrust-All	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
TidalTrust	Original	0.6759	0.5756	1.7259	1.6045	0.7287	0.6368	0.9687	0.7564
	All	0.7432	0.6255	1.5910	1.4226	0.7653	0.6713	0.8355	0.6465
	Interpersonal	0.6833	0.5675	1.6996	1.5169	0.7674	0.6548	0.7929	0.6177
	Impersonal	0.7624	0.6469	1.5873	1.4574	0.7710	0.6787	0.9390	0.7315
Merge	Original	0.7441	0.5920	1.5490	1.3336	0.7601	0.6103	0.8788	0.6919
	All	0.7608	0.6140	1.5295	1.3490	0.7752	0.6324	0.8751	0.6892
	Interpersonal	0.7234	0.5734	1.5224	1.3270	0.7791	0.6384	0.8748	0.6890
	Impersonal	0.7890	0.6336	1.4930	1.3172	0.7811	0.6396	0.8766	0.6904
SocialMF	Original	1.4075	1.2177	–	–	–	–	1.0608	0.7760
	All	1.3910	1.1820	–	–	–	–	0.9950	0.7310
	Interpersonal	1.4455	1.2414	–	–	–	–	1.0639	0.7684
	Impersonal	1.6103	1.3370	–	–	–	–	1.081	0.7917

Effectiveness of the Proposed Model

We predict the trust values based on the learned regression models for three scenarios: “All”, “Interpersonal” and “Impersonal”. “All” refers to considering both interpersonal and impersonal aspects, while the others refer to only considering interpersonal or impersonal aspects respectively. The effectiveness of these aspects in predicting trust values are investigated by applying them in three algorithms (i.e. TidalTrust, Merge1 and SocialMF) in terms of predictive accuracy for recommender systems. Further, the three different views²⁷ mentioned in Section 7.2.3 of datasets are studied. Lastly, we further employ the learned regression models from Epinions1 and Epinions2 to FilmTrust and Flixster where distrust information is unavailable.

Before evaluating our performance, we first present the ratio of “reliable” trust links to the original ones according to Equation 7.12. As illustrated in Figure 7.2, a

²⁷In this work, we did not investigate views of controversial items and niche items for the SocialMF algorithm due to the inconvenience of implementation.

Table 7.6: Performance Comparison Based on Different Trust Aspects Using Epinions2

Methods	Aspects	All		Controversial Items 2,653 users 12,775 ratings		Niche Items 8,922 users 731,116 ratings		Flixster-All	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
TidalTrust	Original	0.6420	0.5287	1.6467	1.5444	0.7346	0.6399	1.2449	0.9846
	All	0.7123	0.5857	1.7390	1.5215	0.7867	0.6764	1.2129	0.9706
	Interpersonal	0.6255	0.5132	1.6607	1.4679	0.7814	0.6627	1.2190	0.9789
	Impersonal	0.7194	0.5951	1.6055	1.4950	0.7766	0.6746	1.2449	0.9846
Merge	Original	0.6801	0.5540	1.4869	1.3235	0.7298	0.5934	1.0376	0.8163
	All	0.7032	0.5880	1.5467	1.3754	0.7432	0.6128	1.0362	0.8150
	Interpersonal	0.6734	0.5541	1.4868	1.2830	0.7503	0.6211	1.0366	0.8155
	Impersonal	0.7341	0.6043	1.4669	1.3235	0.7501	0.6143	1.0376	0.8163
SocialMF	Original	1.2799	1.1094	–	–	–	–	1.3747	1.0440
	All	1.2559	1.0971	–	–	–	–	1.3716	1.0450
	Interpersonal	1.2603	1.0999	–	–	–	–	1.3838	1.0506
	Impersonal	1.4474	1.2079	–	–	–	–	1.3747	1.0440

substantial ratio of the original trust links are filtered out as “unreliable” ones by our model. The results in Tables 7.5 and 7.6 are based on these “reliable” trust links. Later we show whether this difference would lead to a performance improvement of the three recommendation algorithms.

In the view of *All*, Tables 7.5 and 7.6 show that our model could almost achieve the best performance with regard to RMSE and MAE for all three algorithms on the three datasets. Our method could achieve similar results with the original trust values on Epinions and Flixster, but demonstrate significant differences on FilmTrust. This may be explained by the fact that most ratings on Epinions and Flixster datasets are highly skewed. Specifically, the average ratings are 4.6964, 4.6863 (out of 5) and 3.6560 (out of 4) in Epinions1, Epinions2 and Flixster, respectively (see Table 7.3). In contrast, the average rating in FilmTrust is 3.0028 out of 4. The same trends could be observed in the views of *Controversial Items* due to less skewed distributed ratings, where our approach obtains much better performance than that with original trust value. It should be noted for *Niche Items*, the performance of our method is worse than that with the original trust.

Figure 7.3: Performance Comparison of TidalTrust Method by Considering Interpersonal and Impersonal Aspects

Figure 7.4: Performance Comparison of TidalTrust Method in All View On Epinions1

This is mainly because niche items are defined as those which received fewer than 5 ratings. In that case, the problem of data sparsity becomes more serious as we filter out some recommenders who might provide ratings to niche items. This problem could be addressed by predicting more implicit trust links with our model. The improvements on the three methods over those with the original trust are remarkable (around 0.13 in RMSE and 0.18 in MAE at most), as [Koren \(2010\)](#) points out that small improvements in RMSE may lead to significant improvements in real applications.

Interpersonal and Impersonal Aspects: Figure 7.3 presents the performance of the TidalTrust algorithm by considering only interpersonal or impersonal aspects. As demonstrated in Figure 7.3, we can see that the trust derived from interpersonal aspects (i.e. on rating history) is more effective than that from impersonal factors (i.e. on the trust

Figure 7.5: Performance Comparison of TidalTrust Method in Different Views on Epinions1

and distrust networks) in terms of RMSE. However, for controversial items, the algorithms depending on trust values derived from impersonal aspects perform slightly better than those from interpersonal aspects. This is due to the fact that users' ratings of controversial items are quite dissimilar, increasing the difficulty of extracting valuable information for personalized recommendations according to rating history. On the contrary, the impersonal aspects, modeled based on the trust and distrust networks, would not be affected by those controversial ratings. Hence, they might infer more reliable trust and distrust values. Further, we explore the effectiveness of each interpersonal aspect as well as their combinations without considering the impersonal aspects. The results are presented in Table 7.7 and Figures 7.4 and 7.5, where *B*, *C*, *I* and *P* denote benevolence, competence, integrity and predictability, respectively. Hence all the combinations of trust aspects can be represented by concatenating letters. For example, *B-C* refers to the combination of the benevolence and competence. As can be seen in Table 7.7, overall, the performance increases as more aspects are included. We thus could conclude that all the four interpersonal aspects are reasonable and each of them contributes to the success of our trust and distrust prediction.

A clearer and more detailed demonstration is illustrated in Figures 7.4 and 7.5 which present the comparison of different interpersonal aspects in terms of performance gaps in

Table 7.7: Performance Comparison of Tidaltrust Based on Interpersonal Aspects on Flixster

# Aspects	RMSE	MAE
1	1.2415 ± 0.0031	0.9828 ± 0.0036
2	1.2383 ± 0.0055	0.9805 ± 0.0043
3	1.2352 ± 0.0028	0.9783 ± 0.0034
4	1.2129	0.9706

different views of datasets. The histogram under the horizontal solid line (representing the original trust performance) means a better performance than the baseline in terms of RMSE. For single aspect, benevolence achieves the best performance than the other three aspects. In contrast, competence obtains the worse performance than predictability or integrity whose performance is equivalent to that of the original trust values. Hence, although competence is an important aspect for trust modelling (see Table 7.4), it is not that useful in recommender systems. Furthermore, the performance of $B - P$ is better than that of B (the best of single aspect) and that of $B - I - P$ or $B - C - P$ (the best of the combinations of three aspects). This implies that predictability, modeled in a different way and providing additional information, can complement benevolence in building the trust relationship in recommender systems. However, integrating with other aspects (e.g. competence or integrity) may not result in better performance. For the best combination $B - P$, benevolence is closely related to individuals' similarity, and predictability, on the contrary, provides indications of the consistency of the similarity trend. In this sense, the two aspects are complementary to each other, and capable of generating better trust values for recommender systems. However, when other aspects are incorporated, redundant and even noisy information could be brought in, which deteriorates performance.

Generalization: It is observed that similar trends of performance are obtained on FilmTrust and Flixster using the coefficients learned from Epinions1 and Epinions2, respectively. As illustrated in Tables 7.5 and 7.6, TidalTrust, Merge and SocialMF

Figure 7.6: Performance Comparison of TidalTrust Method in All View

Figure 7.7: Performance by Varying the Size of Predicted Trust Network on FilmTrust

could achieve better performance with the trust information learned by using the trained logistic regression models (for trust and distrust) on Epinions. Moreover, as shown in Figures 7.6(a) and 7.6(b), benevolence consistently shows better performance than other single aspects, and the combination of benevolence and predictability reaches the best performance among the overall 15 combinations of impersonal aspects. Hence, we conclude that the trust model learned from one dataset can be applied to other datasets where distrust information is unavailable. It is important because most real-world datasets do not contain such information due to various reasons such as privacy concern. In other words, the knowledge learned from one community can be (partially) reused to model the trust and distrust in other communities.

Figure 7.8: Performance Comparison of TidalTrust between “without” and “with” Distrust

The Size of the Predicted Trust Network: Figure 7.7 shows the performance of the three algorithms on FilmTrust dataset by varying the size of the predicted trust network. Here, the X-axis refers to that the predicted trust network is certain times as large as the original network. As demonstrated in the figure, we can see that all three algorithms obtain better recommendation accuracy as the size of the trust network increases, verifying the effectiveness of our model in predicting implicit trust and distrust values (or relationships). Note that both Merge and SocialMF incorporate the mechanism of trust propagation to improve the recommendation accuracy. Therefore, the corresponding performance of incorporating our model could be further improved if the propagation length is made larger than 1 (especially for the Merge method (Guo et al., 2012)).

Predicted Distrust Information: In our model, we employ the predicted distrust value to refine the trust value according to Equations 7.12 and 7.13. Figure 7.8 illustrates the performance comparison of the TidalTrust algorithm on three datasets by differentiating between considering and not considering the predicted distrust information. As illustrated, we can see that the performance of TidalTrust is slightly improved if predicted implicit distrust information is incorporated into trust value prediction. This demonstrates that the noisy (less reliable) trust links could be validly removed by our model.

7.3 Summary

In Section 7.1, we thoroughly compare our four models (i.e. DiffTrust, SARC, PGTM and SubGroup) from the six perspectives of information availability, subjectivity, dishonesty, dynamic, new users and accuracy. Of the four models, DiffTrust can deal with dishonest advisors, capture advisors' dynamic behavior, cope with new users and also relatively accurately model advisors' trustworthiness value. SARC can accurately model users (advisors)' subjectivity, but only partially address dishonesty and dynamic perspectives. PGTM and SubGroup can handle both the subjectivity and dishonesty perspectives. However, PGTM can also accurately measure advisors' trustworthiness but does not perform so well in the other two perspectives, while SubGroup can capture the dynamic behavior of users (or advisors). We then explore the potential of these four models to resist five representative attacks, including Constant attack, Camouflage attack, Whitewashing attack, Sybil attack and Collusion attack. Our analysis shows that: (1) DiffTrust can be robust against the Constant attack, Camouflage attack and Collusion attack, while SARC is relatively vulnerable to these three types, but defends the Whitewashing and Sybil attacks well; and (2) PGTM is vulnerable to the Camouflage attack, but able to resist the other four attack types, while SubGroup might fail to defend the Collusion attack, but is robust against the others.

Section 7.2 explores the multiple facets of trust and distrust predictions for recommender systems, aiming to fill in the gap between trust and distrust as multi-aspect concepts and the relatively simple use of trust and (especially) distrust in recommender systems. Specifically, we identify both the interpersonal and impersonal aspects according to trust theory from social science. The four interpersonal aspects, namely benevolence, competence, integrity and predictability, are formally defined in the trust theory based on which they are computationally modeled in the light of user experience (i.e. ratings) in the systems, while the impersonal aspects are computed on the basis of users' trust and distrust network. Then the importance of each aspect to trust or distrust is learned by applying a logistic regression model trained by real-world datasets that contain trust or

distrust information.

After learning the two logistic regression models, we predict the (implicit) trust and distrust values, where the trust values are further refined by the distrust values. These newly generated trust values are taken as input to three representative trust-based recommendation algorithms (i.e TidalTrust, Merge and SocialMF) in order to validate the effectiveness of our proposed model. The experimental results show that: (1) benevolence and competence are positively correlated with trust whereas the integrity and predictability are negatively correlated. On the other hand, competence and predictability are positively correlated with distrust whereas the benevolence and integrity are negatively correlated. All the four aspects are useful in that each individual aspect can achieve comparable performance derived from the original trust values; (2) the learned trust models can be applied to other communities where distrust information is not available for evaluating both the trust and distrust relationships. Our results could serve as a guidance to effectively build implicit trust or distrust networks (competitive to robust explicit networks) based on our proposed framework when users had no explicit trust or distrust information. Incorporating distrust information could effectively remove noisy and redundant data in the original explicit trust network; (3) both the interpersonal and impersonal aspects are useful. In addition, the interpersonal aspects would take greater effect when there are lots of rating data of users, whereas the impersonal aspects would contribute more to the controversial items; and (4) our ability to predict implicit trust values could bootstrap the trust network, which could further improve the performance of trust-aware recommender systems.

CHAPTER 8

Conclusions and Future Work

In this chapter, we provide concluding remarks for the dissertation. Particularly, we summarize each study and recap the corresponding contributions. We then identify the limitations of the studies, and also discuss the possible directions for future research.

8.1 Concluding Remarks

In the dissertation we have presented four models to fulfill two major objectives: (1) to solve the subjectivity and dishonesty problems for opinion evaluation; and (2) to simultaneously take social factors (or social theory) into consideration.

First, we propose DiffTrust, a novel trust model stemmed from *diffusion theory* in Social Science, to evaluate the opinions of advisors by modeling their trustworthiness in Chapter 3. This is built on the basis of the following two points. On the one side, trust is recognized as a diffusive concept. When modeling trust, it is crucial to consider the processes through which trust is cultivated in a system. On the other side, diffusion theory 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. In particular, the advisor's trustworthiness perceived by a specific user is influenced by four important factors: 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 spatial and temporal information about the environment. DiffTrust can

capture the dynamics of trust, and its dependency on other users and the environment. Experimental results on four real datasets (i.e. eBay, FilmTrust, Epinions and Flixster) demonstrate that DiffTrust can consistently perform better in both loosely-connected and well-connected environments than TRAVOS, the Personalized approach, CertProp and Shin in terms of MAE, Precision and MCC. Further, it can help model the trustworthiness of advisors for users who are new to the system. DiffTrust mainly deal with the advisors' dishonesty issue. However, it cannot accurately distinguish between subjectivity difference and dishonesty.

Second, in Chapter 4, we design the SARC model, a subjectivity alignment approach for reputation computation when aggregating numerical ratings (opinions) provided by users towards the same entities. Subjectivity of users may come from two sources by analyzing the scenario of a user providing a rating towards an entity, from both psychological and behavioral perspectives: (1) intra-attribute subjectivity, the subjectivity in evaluating the same attribute; and (2) extra-attribute subjectivity, the subjectivity in evaluating different attributes. We learn these two kinds of subjectivity for each user by applying a Bayesian learning approach and regression analysis on the basis of each user's past experiences (ratings and detailed reviews containing a set of attributes to describe each piece of experience), respectively. Ratings provided by one user can thus be aligned for another user according to the two users' subjectivity. Experimental results in a simulated e-commerce environment verify that (1) SARC can more accurately and stably model sellers' reputation; (2) it is capable of coping with environments with dynamic buyer and seller behavior; and (3) its requirement for successful implementation is not very restrictive. Although SARC only considers subjectivity, it is not much affected by dishonest users, as validated in our experiments.

Third, we also propose two approaches by explicitly distinguishing subjectivity and dishonesty. In Chapter 5, we present our PGTM model, a novel probabilistic graphical trust model to separately consider these two aspects. We adapt the well-known trust framework in Social Science, and model the factors of advisors' intrinsic nature

(dishonesty, 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. We capture the relationship between these factors and trust through the chains in the graphical model. We compare our model with BLADE and Prob-Cog on three real datasets (i.e. FilmTrust, Epinions and Flixster). Experimental results indicate that the latent variables in our model are both theoretically reasonable and computationally effective, and dishonesty and subjectivity are successfully distinguished. Further, we demonstrate that our model can more accurately model advisor trustworthiness without using the partially observable trust links. However, PGTM ignores the fact that dishonesty and subjectivity overlap with each other to a certain extent.

In this view, in Chapter 6, we develop the SubGroup model, a two-layered clustering approach that models each advisor as part of groups. Particularly, we come up with several propositions to indicate features for the clustering analysis for distinguishing subjectivity and dishonesty in users' rating behavior. The values of these features are expected to mirror the change of users' behavior over time, and thus the clustering analysis can also tackle users' dynamic and evolving behavior. The clustering analysis consists of two layers. In the first layer, on the basis of those indicative features, we employ DENCLU twice to crisply cluster advisors into different subjectivity groups and dishonest types. In the second layer, we adapt a fuzzy process to softly smooth and justify the clustering results of the first layer. Each advisor is assigned to two groups with respective membership degrees. Moreover, given the clustering results of the advisors (either subjective or dishonest), we further propose a simple alignment approach to help each user align ratings contributed by advisors to those of her own. Experimental results on both a simulated distributed environment and three real datasets (i.e. FilmTrust, Epinions and Flixster, considered as centralized environments), verify that our approach can achieve better performance in comparison with the state-of-the-art trust models and recommender algorithms. They also demonstrate that users can still extract valuable

information from the opinions contributed by dishonest advisors (i.e. direct dishonest and indirect dishonest types). The opinions from misguidance dishonest advisors (whose behavior follows no pattern) are discarded by our approach.

In addition to presenting and rigorously evaluating these four novel models, in Chapter 7, we perform a thorough comparison of our four models from the following six perspectives: information availability, subjectivity, dishonesty, dynamic, new users and accuracy, where the last five are key success factors for an effective trust model. Moreover, we analyze the potential of our models to cope with other challenging scenarios such as resisting various attacks.

Finally, we use recommender systems as an example to demonstrate how social factors in the well-representative trust framework from Social Science can be applied to improve the performance of existing trust-aware recommender systems. We explore the multiple facets of trust and distrust predictions for recommender systems, aiming to fill in the gap between trust and distrust as multi-aspect concepts and the relatively simple use of trust and (especially) distrust in recommender systems. Specifically, we identify both the interpersonal (i.e. benevolence, competence, integrity and predictability in trust framework from Social Science) and impersonal aspects (identified from (dis)trust networks) according to the trust framework. Based on these aspects, we then regenerate the trust and distrust values, which are taken as input to three representative trust-based recommendation algorithms (i.e. TidalTrust, Merge and SocialMF) in our experiments. The experimental results show that: (1) all the four interpersonal aspects are useful in that each individual aspect can achieve comparable performance derived from the original trust values; (2) the learned trust models can be applied to other communities where distrust information is not available for evaluating both the trust and distrust relationships; (3) both the interpersonal and impersonal aspects are useful. In addition, the interpersonal aspects would take greater effect when there were lots of rating data of users, whereas the impersonal aspects would contribute more to the controversial items; and (4) our ability to predict the implicit trust values could bootstrap the trust network, which could further

improve the performance of trust-aware recommender systems.

In summary, as a new attempt to consider social factors in trust assessment, our approaches contribute to bridging the research gap between computational trust in Computer Science and psychological and behavioral trust in Social Science. Further, we hope that these in-depth studies induce more attention towards this important interdisciplinary research direction.

8.2 Future Work

Our current work focuses on modeling the trustworthiness of advisors to solve the problem of opinion quality. Based on these studies, we also propose some potential and interesting future research directions.

8.2.1 Extending the Current Trust Models

For the future, we plan to extend our current trust models as follows.

First, our current DiffTrust model focuses on modeling the trustworthiness of advisors. We will derive a trust model also from diffusion theory to model the trustworthiness (reputation) of entities to which the advisors provide opinions. Together, they will offer a unified trust model for users to make informed decisions about which entities to interact with.

Second, 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.

Third, we will extend our current PGTM model to address other scenarios (e.g. multi-nominal degrees of trust), and further verify its effectiveness in other applications (e.g. recommender systems). We may also combine the PGTM (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 using the PGTM (or DiffTrust) model, and in the second layer, opinions from advisors with subjectivity difference are aligned by adopting the SARC approach.

Finally, for the SubGroup model, we may further explore the rating behavior of each user (or advisor) in each cluster (i.e. either subjectivity or dishonesty) by using different techniques such as visually plotting their ratings over the time (or over the entities in different categories). In this way, we can construct a more detailed understanding and comparison about the behavior of different users (or advisors) in different clusters, and thus to further improve our model.

8.2.2 A Testbed for Evaluating Trust Models

Our models mainly deal with the subjectivity and dishonesty issues in providing opinions. However, due to the lack of ground truth in the real datasets we used (i.e. eBay, Epinions, Flixster and FilmTrust), we have limited knowledge to exactly tell whether a misleading opinion from an advisor to a user is caused by subjectivity difference between the advisor and the user, or dishonesty of the advisor. Alternatively, instead of evaluating an opinion directly, we evaluate our models by checking whether the outcome of an unknown interaction (between a user and an entity) can be successfully predicted using our models on the basis of a set of opinions previously provided by advisors toward the entity. Although via this kind of substitution we can relatively compare the effectiveness of different models, we are still uncertain about some key problems including: “*To what extent are subjectivity and dishonesty successfully distinguished and addressed?*”, and “*Is a misleading opinion caused by subjectivity or dishonesty?*” etc. Further, due to the limitations of the available datasets, we are not able to fully examine some of the components in our models. For example, in the DiffTrust model, we did not consider context information and physically spatial information in our experiments.

Towards the aforementioned problems, we plan to construct a testbed for evaluating

the trust models particularly for addressing subjectivity and dishonesty for opinion evaluation. The testbed is expected to fulfill three major goals: (1) most important of all, to measure the quality of opinions contributed by different advisors; (2) to evaluate to what extent a model has successfully addressed subjectivity, or dishonesty, or both; and (3) to tell whether a misleading opinion is caused by subjectivity difference, or dishonesty. To achieve these goals, as presented in (Irissappane and Zhang, 2014), in our testbed, instead of using real datasets, we will create the environments using simulations. In particular, we will define the rating behavior of each user by exactly identifying some rating patterns related to subjectivity or dishonesty (Fang et al., 2014; Feng et al., 2012a). In this case, we can define some quantitative measurements about the subjectivity and dishonesty of each rating, and each user (or advisor). Further, in addition to ratings, we will incorporate other features to describe each interaction between a user and an entity to construct more information that might be possibly adopted by trust models, such as context information in DiffTrust (Fang et al., 2013c), and objective attributes in SARC (Fang et al., 2012b).

8.2.3 Checking the Robustness of Trust Models in Resisting Attacks

To improve the robustness of trust models in resisting various attacks has increasingly attracted researchers' attention in the computational trust area (Jøsang and Golbeck, 2009; Zhang et al., 2012; Jøsang, 2012; Jiang et al., 2013). In this case, an interesting future direction for our research is to quantitatively analyze how our trust models can effectively defend various attacks.

Following the quantitative definitions of attacks in existing studies of (Zhang et al., 2012; Jiang et al., 2013), we aim to resolve the following research questions: (1) to what extent each social factor (e.g. benevolence and competence) in our trust models can contribute to the success of resisting each attack (e.g. sybil and whitewashing attack); (2) to what extent considering subjectivity, or dishonesty, or both in our trust models can help successfully defend each attack. All the experiments can be conducted in the testbed

outlined in Section 8.2.2. Based on the experimental results, we can further design more robust trust models by concerning social factors, and dishonesty and subjectivity issues, and also come up with more suitable robustness measurements for our research scenarios.

8.2.4 Merging Review Evaluation with Rating Evaluation

Our four studies mainly focus on evaluating the trustworthiness of advisors by using only *ratings* but overlooking other information such as *textual reviews*. However, in online communities, opinions mostly compose of both the numerical ratings and textual reviews, which are regarded to involve far richer information than numerical ratings. It is thus worthwhile to incorporate textual reviews in optimizing trust models to further improve the effectiveness of trust models in measuring the trustworthiness of advisors (i.e. reviewers) ([Şensoy et al., 2013](#)), and then more accurately measure the quality of opinions.

Merging review evaluation with rating evaluation for better trust models is an interesting direction for future research, especially in the computational trust area. The problem of detecting spam reviews (i.e. reviews involving untruthful information) has been widely studied by researchers by using machine learning and data mining techniques ([Li et al., 2011](#); [Xu et al., 2013](#)). A combination of these existing approaches with trust models may prove fruitful: (1) the analysis results from those models for detecting spam reviews can be adopted to revise the trust values in our trust models; (2) the linguistic features ([Ott et al., 2011](#)) extracted from reviews, together with the numerical features from ratings, can be applied to better distinguish dishonesty and subjectivity from advisors' behavior, and then employed by current trust models (e.g. SubGroup). In addition, linguistic features can be associated with social factors (e.g. benevolence and integrity) in the trust framework from Social Science; and (3) ratings can be merged from the point of view of information uncertainty in information theory ([Klir, 2005](#)) to construct better trust models.

8.2.5 Incorporating Other Suitable Social Theories

The research in this dissertation demonstrates that social theories and frameworks in Social Science can be successfully adopted by computational trust models. In the future, other social theories could be considered. Social dynamics refers to “the behavior of groups that results from the interactions of individual group members as well to the study of the relationship between individual interactions and group level behavior” ([Durlauf and Young, 2004](#)). It has two leading characteristics: (1) it is concerned with changes over time and emphasizes the role of feedback; and (2) each individual is influenced by one another’s behavior. In this view, we can clearly see the similarity between their research subjects with trust, since an advisor’s trustworthiness is dynamic, and will react to the environment and be influenced by others.

Further, social cognitive theory ([Bandura, 2001](#)) in Psychology is a possible candidate. It provides a framework for understanding, predicting, and changing human behavior, and identifies human behavior as an interaction of personal factors, behavior, and the environment. In social cognitive theory, interactions occur between each pair of the three parties: environment (e.g. online communities), user, and behavior (e.g. rating behavior). Interaction between user and behavior involves the influence of the user’s behavior, while interaction between user and environment involves human beliefs and cognitive competencies that are developed and modified by social influence and structures within the environment. The third one, between environment and behavior, involves user’s behavior determining the aspects of their environment and in turn their behavior is modified by that environment. Based on this theory, we can formalize components in modeling advisors’ trustworthiness.

References

- Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749.
- Angst, C. M., Agarwal, R., Sambamurthy, V., and Kelley, K. (2010). Social contagion and information technology diffusion: The adoption of electronic medical records in U.S. hospitals. *Management Science*, 56(8):1219–1241.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52(1):1–26.
- Bickart, B. and Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, 15(3):31–40.
- Biyani, P., Bhatia, S., Caragea, C., and Mitra, P. (2012). Thread specific features are helpful for identifying subjectivity orientation of online forum threads. In *Proceeding of the 24th International Conference on Computational Linguistics (COLING)*, pages 295–310.
- Bizer, C., Heath, T., and Berners-Lee, T. (2009). Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22.
- Blair, J. P., Levine, T. R., and Shaw, A. S. (2010). Content in context improves deception detection accuracy. *Human Communication Research*, 36(3):423–442.
- Bond Jr, C. F., Howard, A. R., Hutchison, J. L., and Masip, J. (2013). Overlooking the obvious: Incentives to lie. *Basic and Applied Social Psychology*, 35(2):212–221.

- Casella, G. and George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174.
- Chai, T. and Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae)?—arguments against avoiding rmse in the literature. *Geoscientific Model Development*, 7:1247–1250.
- Chen, P.-Y., Dhanasobhon, S., and Smith, M. D. (2008). All reviews are not created equal: The disaggregate impact of reviews and reviewers at amazon.com. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=918083.
- Chevalier, J. and Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 8:345–354.
- Chirita, P.-A., Nejdl, W., and Zamfir, C. (2005). Preventing shilling attacks in online recommender systems. In *Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management (WIDM)*, pages 67–74. ACM.
- Chowdhury, M., Thomo, A., and Wadge, B. (2009). Trust-based infinitesimals for enhanced collaborative filtering. In *Proceedings of the 15th International Conference on Management of Data (COMAD)*.
- Chua, F. C. T. and Lim, E.-P. (2010). Trust network inference for online rating data using generative models. In *Proceedings of the 16th ACM Conference on Knowledge Discovery and Data Mining (SIGKDD)*, pages 889–898.
- Coetzee, D., Fox, A., Hearst, M. A., and Hartmann, B. (2014). Should your mooc forum use a reputation system? In *Proceedings of the 17th ACM conference on Computer Supported Cooperative Work & Social Computing (CSCW)*, pages 1176–1187. ACM.
- Dellarocas, C. (2005). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10):1407–1424.

- Dellarocas, C., Zhang, X. M., and Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive marketing*, 21(4):23–45.
- Delli Carpini, M. X. (2000). Gen. com: Youth, civic engagement, and the new information environment. *Political Communication*, 17(4):341–349.
- Doane, D. P. and Seward, L. E. (2011). Measuring skewness: a forgotten statistic. *Journal of Statistics Education*, 19(2):1–18.
- Doh, S. J. and Hwang, J. S. (2005). How consumers evaluate ewom (electronic word-of-mouth) messages. *Cyber Psychology & Behavior*, 12(2):56–72.
- Durlauf, S. N. and Young, H. P. (2004). *Social dynamics*, volume 4. Mit Press.
- Etuk, A., Norman, T. J., Sensoy, M., Bisdikian, C., and Srivatsa, M. (2013). Tidy: A trust-based approach to information fusion through diversity. In *Proceedings of the 16th International Conference on Information Fusion (FUSION)*, pages 1188–1195.
- Falcone, R. and Castelfranchi, C. (2001). Trust and deception in virtual societies. chapter Social trust: a cognitive approach, pages 55–90. Kluwer Academic Publishers.
- Fang, H., Bao, Y., and Zhang, J. (2013a). Misleading opinions provided by advisors: Dishonesty or subjectivity. In *Proceeding of the 23rd International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 1983–1989.
- Fang, H., Zhang, J., Bao, Y., and Zhu, Q. (2013b). Towards effective online review systems in the chinese context: A cross-cultural empirical study. *Electronic Commerce Research and Applications*, 12(3):208–220.
- Fang, H., Zhang, J., and Magnenat-Thalmann, N. (2013c). A trust model stemmed from the diffusion theory for opinion evaluation. In *Proceedings of the 12th International Conference on Autonomous Agents and Multi-agent Systems (AAMAS)*, pages 805–812. International Foundation for Autonomous Agents and Multiagent Systems.

- Fang, H., Zhang, J., and Magnenat-Thalmann, N. (2014). Subjectivity grouping: learning from users' rating behavior. In *Proceedings of the 13th International Conference on Autonomous Agents and Multi-agent Systems (AAMAS)*, pages 1241–1248. International Foundation for Autonomous Agents and Multiagent Systems.
- Fang, H., Zhang, J., Sensoy, M., and Magnenat-Thalmann, N. (2012a). A generalized stereotypical trust model. In *Proceedings of 11th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pages 698–705.
- Fang, H., Zhang, J., Sensoy, M., and Magnenat-Thalmann, N. (2012b). SARC: subjectivity alignment for reputation computation. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1365–1366.
- Feng, Q., Liu, L., and Dai, Y. (2012a). Vulnerabilities and countermeasures in context-aware social rating services. *ACM Transactions on Internet Technology (TOIT)*, 11(3):11.
- Feng, S., Xing, L., Gogar, A., and Choi, Y. (2012b). Distributional footprints of deceptive product reviews. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 98–105.
- Figge, S. (2004). Situation-dependent services - a challenge for mobile network operators. *Journal of Business Research*, 57(12):1416–1422.
- Garcin, F., Xia, L., and Faltings, B. (2013). How aggregators influence human rater behavior? In *Proceedings of the Workshop on International Conference on Electronic Commerce (ICEC)*.
- George, T. and Merugu, S. (2005). A scalable collaborative filtering framework based on co-clustering. In *Proceedings of the 5th IEEE International Conference on Data Mining (ICDM)*, pages 4–pp.

- Ghose, A. and Ipeirotis, P. G. (2007). Designing novel review ranking systems: predicting the usefulness and impact of reviews. In *Proceedings of the 9th International Conference on Electronic Commerce (ICEC)*, pages 303–310. ACM.
- Ghose, A. and Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10):1498–1512.
- Golbeck, J. (2006). Generating predictive movie recommendations from trust in social networks. In *Trust Management*, pages 93–104.
- Golbeck, J. A. (2005). *Computing and applying trust in web-based social networks*. PhD thesis, University of Maryland.
- Gunes, I., Kaleli, C., Bilge, A., and Polat, H. (2012). Shilling attacks against recommender systems: a comprehensive survey. *Artificial Intelligence Review*, pages 1–33.
- Guo, G., Zhang, J., and Thalmann, D. (2012). A simple but effective method to incorporate trusted neighbors in recommender systems. In *Proceedings of the 20th Conference on User Modeling, Adaption, and Personalization (UMAP)*, pages 114–125.
- Guo, G., Zhang, J., and Yorke-Smith, N. (2013). A novel bayesian similarity measure for recommender systems. In *Proceedings of the 23rd International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 2619–2625.
- Gurun, U. G. and Butler, A. W. (2012). Don’t believe the hype: local media slant, local advertising, and firm value. *The Journal of Finance*, 67(2):561–598.
- Haghpanah, Y. and desJardins, M. (2012). PRep: a probabilistic reputation model for biased societies. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 315–322.
- Hang, C.-W., Wang, Y., and Singh, M. P. (2009). Operators for propagating trust and their

- evaluation in social networks. In *Proceedings of the 8th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 1025–1032.
- Hang, C.-W., Zhang, Z., and Singh, M. P. (2013). Shin: Generalized trust propagation with limited evidence. *IEEE Computer*, 46(3):78–85.
- Harrell, F. E., Lee, K. L., Califf, R. M., Pryor, D. B., and Rosati, R. A. (1984). Regression modelling strategies for improved prognostic prediction. *Statistics in Medicine*, 3(2):143–152.
- Hennig-Thurau, T., Gwinner, K. P., and Walsh, G. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet. *Journal of Interactive Marketing*, 18(1).
- Hinneburg, A. and Keim, D. A. (1998). An efficient approach to clustering in large multimedia databases with noise. In *Proceedings of the 4th ACM Conference on Knowledge Discovery and Data Mining (SIGKDD)*, pages 58–65.
- Hosmer, L. T. (1995). Trust: The connecting link between organizational theory and philosophical ethics. *Academy of Management Review*, 20(2):379–403.
- Hu, N., Bose, I., Gao, Y., and Liu, L. (2011a). Manipulation in digital word-of-mouth: A reality check for book reviews. *Decision Support Systems*, 50(3):627–635.
- Hu, N., Bose, I., Koh, N. S., and Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, 52(3):674–684.
- Hu, N., Liu, L., and Sambamurthy, V. (2011b). Fraud detection in online consumer reviews. *Decision Support Systems*, 50:614–626.
- Hu, N., Liu, L., and Zhang, J. J. (2008). Do online reviews affect product sales? the role of reviewer characteristics and temporal effects. *Information Technology and Management*, 9(3):201–214.

- Huynh, T. D., Jennings, N. R., and Shadbolt, N. R. (2006). An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 13(2):119–154.
- Irissappane, A. A., Jiang, S., and Zhang, J. (2014a). A biclustering-based approach to filter dishonest advisors in multi-criteria e-marketplaces. In *Proceedings of the 13th International Conference on Autonomous Agents and Multi-agent Systems (AAMAS)*, pages 1385–1386. International Foundation for Autonomous Agents and Multiagent Systems.
- Irissappane, A. A., Oliehoek, F. A., and Zhang, J. (2014b). A POMDP based approach to optimally select sellers in electronic marketplaces. pages 1329–1336.
- Irissappane, A. A. and Zhang, J. (2014). A testbed to evaluate the robustness of reputation systems in e-marketplaces. In *Proceedings of the 13th International Conference on Autonomous Agents and Multi-agent Systems (AAMAS)*, pages 1629–1630. International Foundation for Autonomous Agents and Multiagent Systems.
- Jamali, M. and Ester, M. (2009). Trustwalker: a random walk model for combining trust-based and item-based recommendation. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 397–406. ACM.
- Jamali, M. and Ester, M. (2010). A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the 4th ACM Recommender Systems Conference (RecSys)*, pages 135–142.
- Ji, A.-T., Yeon, C., Kim, H.-N., and Jo, G.-S. (2007). Distributed collaborative filtering for robust recommendations against shilling attacks. In *Advances in Artificial Intelligence*, pages 14–25. Springer.
- Jiang, S., Zhang, J., and Ong, Y.-S. (2013). An evolutionary model for constructing robust

- trust networks. In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 813–820.
- Jindal, N. and Liu, B. (2008). Opinion spam and analysis. In *Proceedings of the 1st International Conference on Web Search and Data Mining (WSDM)*, pages 219–230.
- Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, 37:183–233.
- Jøsang, A. (2012). Robustness of trust and reputation systems: Does it matter? In *Trust Management VI*, pages 253–262. Springer.
- Jøsang, A. and Golbeck, J. (2009). Challenges for robust trust and reputation systems. In *Proceedings of the 5th International Workshop on Security and Trust Management (SMT)*, pages 1–12.
- Jøsang, A. and Ismail, R. (2002). The beta reputation system. In *Proceedings of the 15th Bled Electronic Commerce Conference*, pages 41–55.
- Jøsang, A., Ismail, R., and Boyd, C. (2007). A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644.
- Jøsang, A., Quattrociocchi, W., and Karabeg, D. (2011). Taste and trust. In *Trust Management*, pages 312–322.
- Kerr, R. and Cohen, R. (2009a). An experimental testbed for evaluation of trust and reputation systems. In *Trust Management III*, pages 252–266. Springer.
- Kerr, R. and Cohen, R. (2009b). Smart cheaters do prosper: defeating trust and reputation systems. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 993–1000.
- Klir, G. J. (2005). *Uncertainty and information: foundations of generalized information theory*. John Wiley & Sons.

- Koren, Y. (2010). Factor in the neighbors: Scalable and accurate collaborative filtering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 4(1):1–24.
- Korfiatis, N., García-Bariocanal, E., and Sánchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, 11(3):205–217.
- Koster, A., Sabater-Mir, J., and Schorlemmer, M. (2010). Inductively generated trust alignments based on shared interactions. In *Proceedings of 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1571–1572.
- Kwon, K., Cho, J., and Park, Y. (2009). Multidimensional credibility model for neighbor selection in collaborative recommendation. *Expert Systems with Applications*, 36(3):7114–7122.
- Lam, S. K. and Riedl, J. (2004). Shilling recommender systems for fun and profit. In *Proceedings of the 13th International Conference on World Wide Web (WWW)*, pages 393–402. ACM.
- Lauw, H. W., Lim, E.-P., and Wang, K. (2008). Bias and controversy in evaluation systems. *IEEE Transactions on Knowledge and Data Engineering*, 20(11):1490–1504.
- Lee, J., Park, D., and Han, I. (2008). The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications*, 7:341–352.
- Lee, J., Park, D. H., and Han, I. (2011). The different effects of online consumer reviews on consumers’ purchase intentions depending on trust in online shopping malls: An advertising perspective. *Internet Research*, 21(2):187–206.
- Lerman, K., Intagorn, S., Kang, J.-H., and Ghosh, R. (2012). Using proximity to predict activity in social networks. In *Proceedings of the 21st International Conference Companion on World Wide Web (WWW)*, pages 555–556.

- Levine, T. R., Kim, R. K., and Hamel, L. M. (2010). People lie for a reason: Three experiments documenting the principle of veracity. *Communication Research Reports*, 27(4):271–285.
- Li, F., Huang, M., Yang, Y., and Zhu, X. (2011). Learning to identify review spam. In *Proceedings of the 22nd International Joint cConference on Artificial Intelligence (IJCAI)*, pages 2488–2493. AAAI Press.
- Liu, B. (2010). Sentiment analysis and subjectivity. In Indurkha, N. and Damerau, F. J., editors, *Handbook of natural language processing*, pages 627–666. Chapman & Hall, 2nd edition.
- Liu, J. S. (1994). The collapsed gibbs sampler in bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966.
- Liu, S., Zhang, J., Miao, C., Theng, Y.-L., and Kot, A. C. (2011). iCLUB: An integrated clustering-based approach to improve the robustness of reputation systems. In *Proceedings of the 10th International Conference on Autonomous Agents and Multi-agent Systems (AAMAS)-Volume 3*, pages 1151–1152. International Foundation for Autonomous Agents and Multiagent Systems.
- Liu, S., Zhang, J., Miao, C., Theng, Y.-L., and Kot, A. C. (2014). An integrated clustering-based approach to filtering unfair multi-nominal testimonies. *Computational Intelligence*, 30(2):316–341.
- Liu, X., Datta, A., Fang, H., and Zhang, J. (2012). Detecting imprudence of ‘reliable’sellers in online auction sites. In *Proceedings of the 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pages 246–253.
- Ma, H., Lyu, M. R., and King, I. (2009). Learning to recommend with trust and distrust

- relationships. In *Proceedings of the third ACM Conference on Recommender Systems (RecSys)*, pages 189–196. ACM.
- Ma, H., Yang, H., Lyu, M. R., and King, I. (2008). Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and Knowledge Management (CIKM)*, pages 931–940. ACM.
- Majumdar, S., Kulkarni, D., and Ravishankar, C. V. (2007). Addressing click fraud in content delivery systems. In *Proceedings of IEEE International Conference on Computer Communications (INFOCOM)*, pages 240–248.
- Marsh, S. P. (1994). *Formalising trust as a computational concept*. PhD thesis, Department of Mathematics and Computer Science, University of Stirling.
- Massa, P. and Avesani, P. (2007). Trust-aware recommender systems. In *Proceedings of the 2007 ACM Conference on Recommender Systems (Recsys)*, pages 17–24.
- Mayer, R. C., Davis, J. H., and Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, 20(3):709–734.
- McKnight, D. H. and Chervany, N. L. (2001). What trust means in e-commerce customer relationships: a interdisciplinary conceptual typology. *International Journal of Electronic Commerce*, 6(2):35–59.
- McKnight, D. H. and Choudhury, V. (2006). Distrust and trust in b2c e-commerce: Do they differ? In *Proceedings of the 8th International Conference on Electronic Commerce (ICEC)*, pages 482–491. ACM.
- Michael, B., Wrazien, S., and R., G. (2010). Using machine learning to augment collaborative filtering of community discussions. In *Proceedings of the 9th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1569–1570.

- Mnih, A. and Salakhutdinov, R. (2007). Probabilistic matrix factorization. In *Proceeding of the 21st Conference on Neural Information Processing Systems (NIPS)*, pages 1257–1264.
- Muchnik, L., Aral, S., and Taylor, S. J. (2013). Social influence bias: A randomized experiment. *Science*, 341(6146):647–651.
- Noorian, Z., Marsh, S., and Fleming, M. (2011). Multi-layer cognitive filtering by behavioral modelling. In *Proceedings of the 10th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 871–878.
- O’ Mahony, M. P., Hurley, N. J., and Silvestre, G. C. (2002a). Promoting recommendations: An attack on collaborative filtering. In *Database and Expert Systems Applications (DESA)*, pages 494–503. Springer.
- O’ Mahony, M. P., Hurley, N. J., and Silvestre, G. C. (2002b). Towards robust collaborative filtering. In *Artificial Intelligence and Cognitive Science*, pages 87–94. Springer.
- O’ Mahony, M. P., Hurley, N. J., and Silvestre, G. C. (2003). Collaborative filtering—safe and sound? In *Foundations of Intelligent Systems*, pages 506–510. Springer.
- O’Donovan, J. and Smyth, B. (2005). Trust in recommender systems. In *Proceedings of the 10th International Conference on Intelligent User Interfaces (IUI)*, pages 167–174.
- Oliehoek, F. A., Gokhale, A. A., and Zhang, J. (2012). Reasoning about advisors for seller selection in e-marketplaces via pomdps. In *Proceedings of the 11 International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS) Workshop on Trust in Agent Societies (TRUST)*.
- Opsahl, T., Agneessens, F., and Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251.

- Ott, M., Choi, Y., Cardie, C., and Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 309–319. Association for Computational Linguistics.
- Park, D. H. and Kim, S. (2008). The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. *Electronic Commerce Research and Applications*, 7:399–400.
- Park, S.-t., Chu, W., Beaupre, T., Agarwal, D. K., Roy, S., and Ramakrishnan, R. (2012). Determining user preference of items based on user ratings and user features. US Patent 8,301,624.
- Pearce, K. E. and Rice, R. E. (2013). Digital divides from access to activities: comparing mobile and personal computer internet users. *Journal of Communication*, 63(4):721–744.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo CA.
- Picazo-Vela, S., Chou, S. Y., Melcher, A. J., and Pearson, J. M. (2010). Why provide an online review? an extended theory of planned behavior and the role of big-five personality traits. *Computers in Human Behavior*, 26(4):685–696.
- Pinyol, I., Sabater-Mir, J., Dellunde, P., and Paolucci, M. (2012). Reputation-based decisions for logic-based cognitive agents. *Autonomous Agent Multi-Agent System*, 24:175–216.
- Radinsky, K., Davidovich, S., and Markovitch, S. (2012). Learning causality for news events prediction. In *Proceedings of the 21st International Conference on World Wide Web (WWW)*, pages 909–918.

- Ray, S. and Mahanti, A. (2010). Improving prediction accuracy in trust-aware recommender systems. In *Proceedings of the 43rd Hawaii International Conference on System Sciences (HICSS)*, pages 1–9.
- Regan, K., Poupart, P., and Cohen, R. (2006). Bayesian reputation modeling in E-marketplaces sensitive to subjectivity, deception and change. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*, pages 1206–1212, Boston, Massachusetts, USA. AAAI Press.
- Resnick, P. and Zeckhauser, R. (2002). Trust among strangers in internet transactions: Empirical analysis of ebay’s reputation system. *The Economics of the Internet and E-Commerce*, 11:127–157.
- Resnick, P., Zeckhauser, R., Friedman, E., and Kuwabara, K. (2000). Reputation systems. *Advances in Applied Microeconomics*, 43(12):45–48.
- Resnick, P., Zeckhauser, R., Swanson, J., and Lockwood, K. (2006). The value of reputation on ebay: A controlled experiment. *Experimental Economics*, 9(2):79–101.
- Rogers, E. M., editor (1995). *Diffusion of Innovations*, page 5. Simon and Schuster, 4th edition.
- Sabater, J. and Sierra, C. (2001). Regret: a reputation model for gregarious societies. In *Proceedings of the Fourth Workshop on Deception, Fraud and Trust in Agent Societies*, pages 61–69.
- Sabater, J. and Sierra, C. (2002). Reputation and social network analysis in multi-agent systems. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 475–482.
- Şensoy, M., Fokoue, A., Pan, J. Z., Norman, T. J., Tang, Y., Oren, N., and Sycara, K. (2013). Reasoning about uncertain information and conflict resolution through trust revision. In *Proceedings of the 12th International Conference on Autonomous*

- Agents and Multi-agent Systems (AAMAS)*, pages 837–844. International Foundation for Autonomous Agents and Multiagent Systems.
- Şensoy, M., Zhang, J., Yolum, P., and Cohen, R. (2009). POYRAZ: Context-aware service selection under deception. *Computational Intelligence*, 25(4):335–366.
- Shafer, G. (1976). *A mathematical theory of evidence*, volume 1. Princeton university press Princeton.
- Sheline, Y. I., Sanghavi, M., Mintun, M. A., and Gado, M. H. (1999). Depression duration but not age predicts hippocampal volume loss in medically healthy women with recurrent major depression. *The Journal of Neuroscience*, 19(12):5034–5043.
- Steffes, E. M. and Burgee, L. E. (2009). Social ties and online word of mouth. *Internet Research*, 19(1):42–59.
- Strang, D. and Soule, S. A. (1998). Diffusion in organizations and social movements: from hybrid corn to poison pills. *Annual Review of Sociology*, 24:265–290.
- Strang, D. and Tuma, N. B. (1993). Spatial and temporal heterogeneity in diffusion. *American Journal of Sociology*, 99(3):614–639.
- Swamynathan, G., Almeroth, K. C., and Zhao, B. Y. (2010). The design of a reliable reputation system. *Electronic Commerce Research*, 10(3-4):239–270.
- Teacy, W. T. L., Luck, M., Rogers, A., and Jennings, N. R. (2012). An efficient and versatile approach to trust and reputation using hierarchical bayesian modelling. *Artificial Intelligence*, 193:149–185.
- Teacy, W. T. L., Patel, J., Jennings, N. R., and Luck, M. (2006). TRAVOS: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12(2):183–198.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In

-
- Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 178–185.
- Urbano, J., Rocha, A. P., and Oliveira, E. (2009). Computing confidence values: Does trust dynamics matter? In *Progress in Artificial Intelligence*, pages 520–531. Springer.
- Urbano, J., Rocha, A. P., and Oliveira, E. (2010). In the search of better deals using trust. In *Proceedings of the Workshop on Intelligent Agents and Technologies for e-Business (IAT4EB)*, pages 25–30.
- Utz, S., Kerkhof, P., and van den Bos, J. (2012). Consumers rule: How consumer reviews influence perceived trustworthiness of online stores. *Electronic Commerce Research and Applications*, 11(1):49–58.
- Valente, T. W. (2005). Network models and methods for studying the diffusion of innovations. In Carrington, P. J., Scott, J., and Wasserman, S., editors, *Models And Methods In Social Network Analysis*, pages 98–116. Cambridge University Press.
- Van Eck, P. S., Jager, W., and Leeftang, P. S. (2011). Opinion leaders’ role in innovation diffusion: A simulation study. *Journal of Product Innovation Management*, 28(2):187–203.
- Victor, P., Cornelis, C., De Cock, M., and Teredesai, A. (2011). Trust-and distrust-based recommendations for controversial reviews. *IEEE Intelligent Systems*, 26(1):48–55.
- Victor, P., Verbiest, N., Cornelis, C., and Cock, M. D. (2013). Enhancing the trust-based recommendation process with explicit distrust. *ACM Transactions on the Web (TWEB)*, 7(2):42–59.
- Whitby, A., Jøang, A., and Indulska, J. (2004). Filtering out unfair ratings in bayesian reputation systems. In *Proceedings of the 3rd International Joint Conference on Autonomous Agent Systems Workshop on Trust in Agent Societies (TRUST)*, pages 106–117.

- Williamson, O. E. (1993). Calculativeness, trust, and economic organization. *Journal of Law and Economics*, 36(1):453–486.
- Xu, C., Zhang, J., Chang, K., and Long, C. (2013). Uncovering collusive spammers in chinese review websites. In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM)*, pages 979–988.
- Yang, Y.-F., Feng, Q.-Y., Sun, Y. L., and Dai, Y.-F. (2009). Dishonest behaviors in online rating systems: cyber competition, attack models, and attack generator. *Journal of Computer Science and Technology*, 24(5):855–867.
- Ye, Q., Zhang, Z. Q., and Luo, Z. X. (2007). Research on chinese subjectivity automatically analysis towards emotional tendency of online reviews. *China Journal of Information Systems*, 1(1):79–91.
- Yu, B. and Singh, M. P. (2003a). Detecting deception in reputation management. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 73–80.
- Yu, B. and Singh, M. P. (2003b). Searching social networks. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 65–72.
- Zacharia, G., Moukas, A., and Maes, P. (2000). Collaborative reputation mechanisms for electronic marketplaces. *Decision Support Systems*, 29(4):371–388.
- Zhang, F.-g. and Xu, S.-h. (2007). Analysis of trust-based e-commerce recommender systems under recommendation attacks. In *Proceedings of the 1st IEEE International Symposium on Data, Privacy, and E-Commerce (ISDPE)*, pages 385–390. IEEE.
- Zhang, H., Jiang, L., and Su, J. (2005). Hidden naive bayes. In *Proceedings of the 20th National Conference on Artificial Intelligence (AAAI)*, pages 919–924.

- Zhang, J. and Cohen, R. (2006). Trusting advice from other buyers in e-marketplaces: the problem of unfair ratings. In *Proceedings of the 8th International Conference on Electronic commerce (ICEC)*, pages 225–234. ACM.
- Zhang, J. and Cohen, R. (2007). A comprehensive approach for sharing semantic web trust ratings. *Computational Intelligence*, 23(3):302–319.
- Zhang, J. and Cohen, R. (2008). Evaluating the trustworthiness of advice about selling agents in e-marketplaces: A personalized approach. *Electronic Commerce Research and Applications*, 7(3):330–340.
- Zhang, L., Fang, H., Ng, W. K., and Zhang, J. (2011). Inrank: Interaction ranking-based trustworthy friend recommendation. In *Proceedings of the 10th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (Trust-Com)*, pages 266–273. IEEE.
- Zhang, L., Jiang, S., Zhang, J., and Ng, W. K. (2012). Robustness of trust models and combinations for handling unfair ratings. In *Trust Management VI*, pages 36–51. Springer.
- Zhang, Z. Q., Ye, Q., Law, R., and Li, Y. J. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, 29:694–700.