Modeling witness trustworthiness for reliable reputation systems

Liu, Siyuan

2013

Liu, S. (2013). Modeling witness trustworthiness for reliable reputation systems. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/52235>

<https://doi.org/10.32657/10356/52235>

MODELING WITNESS TRUSTWORTHINESS FOR RELIABLE REPUTATION SYSTEMS

LIU SIYUAN

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2013

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

.....

Date

.....

LIU SIYUAN

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Alex C. Kot, for his continuous guidance and support through my Ph.D study and research. It has been my honor to be one of his Ph.D students. His inspiring nature, patience, immense knowledge, and encouragement have been great source of motivation throughout my doctoral studies. His confidence in my abilities to carry out the research work is greatly appreciated. His emphasis on the demonstrability of the ideas has immensely influenced my research work. His sharp insights and excellent guidance, his encouragements in every stage of my research work, his patience and endeavor in improving every single word in papers, and his criticism and strict attitude towards the research work have been a resource and model for me to complete this thesis and they will continuously benefit me throughout my lifetime.

I would also like to express my gratitude to my other two advisors, Professor Theng Yin-Leng and Miao Chunyan. Their creative research ideas and insightful advice greatly help my research. I wish to express my thanks to Dr. Zhang Jie, Dr. Sun Yan, Dr. Liu Yuhong, and Mr. Yu Han for their enlightening discussions and collaborations. I wish to acknowledge all the panel members of this thesis for their time, valuable comments and supports. I sincerely thank Dr. Yang Huijuan, Dr. Cao Hong, Mr. Chen Changsheng, Mr. Li Sheng, Ms. Fan Jiayuan, Ms. Wang Yan, Mr. Miao Zhenwei, and Mr. Lai Jian from workstation resource lab for their encouragements. I am grateful to Stewart Chu for his kind technical supports.

My dream towards a Ph.D can never be possible without the strong support from my family. I am indebted to my parents (Liu Li and Yuan Ping) for instilling me the perseverance and hard-working from childhood which benefit me no matter what I choose to do. I would like to dedicate this work to my husband Qu Lei for his persistent love, endless support and sacrifice.

Abstract

Reputation systems have been an important part for the success of online service provision systems, where a service can be, for example, an e-commerce transaction or a functional component implemented by Web service technologies. However, the existence of unfair testimonies seriously jeopardizes the reliability of reputation systems. In this thesis, we propose three approaches to address the problem of unfair testimonies. First, we propose a two-stage clustering filtering approach to filter unfair testimonies. Then we propose an integrated clustering filtering approach as an improvement of the two-stage clustering filter approach. Using a different mechanism, we further propose a novel witness trustworthiness model based on Dempster-Shafer Theory. Experimental results show that the proposed approaches can effectively mitigate the influence of unfair testimonies for reputation systems, especially for the ones supporting multi-nominal ratings, counter collusion attacks to a good extent, and eventually contribute to building reliable reputation systems.

Contents

Acknowledgements	i
Abstract	ii
List of Figures	viii
List of Tables	xi
1 Introduction	1
1.1 Motivation	3
1.2 Contributions	7
1.3 Organization of the Thesis	8
2 Literature Review	10
2.1 Approaches to Address the Problem of Unfair Testimonies	10
2.1.1 Iterated Filtering Approach	10
2.1.2 Entropy-Based Filtering Approach	13
2.1.3 Clustering Filtering Approach	14
2.1.4 Two-Layer Cognitive Filtering Approach	15

2.1.5	TRAVOS	17
2.1.6	Yu and Singh's Approach	19
2.1.7	Bayesian Learning Approach	21
2.1.8	CHI-Score Approach	23
2.1.9	Personalized Approach	24
2.1.10	Trust-Based Incentive Approach	26
2.2	Categorization of the Approaches	27
2.3	Capabilities	30
2.4	Factors Impacting the Approaches	33
2.4.1	System Architecture	33
2.4.2	Rating Level Granularity	34
2.5	Summary	35
3	A Two-Stage Clustering Filtering Approach	37
3.1	Introduction	37
3.2	Notations	39
3.2.1	Aggregating Consumers' Ratings	39
3.2.2	Aggregating Ratings with Aging	40
3.3	The Proposed Two-Stage Clustering Filtering Approach	41
3.4	Experimental Studies	45
3.4.1	Threshold Value Exploration	47
3.4.2	Robustness Investigation	53
3.4.2.1	Dishonest Witness Percentage and Transaction Number Varying	53

3.4.2.2	Provider Behavior Varying	55
3.4.3	Comparative Experiments	56
3.4.4	Discussions on Experimental Results	60
3.5	Summary	61
4	An Integrated Clustering Filtering Approach	63
4.1	Introduction	63
4.2	Notations	64
4.3	The iCLUB Approach	65
4.3.1	Making Use of Local Information	66
4.3.2	Making Use of Global Information	67
4.3.3	The Integrated Approach	69
4.4	Examples	70
4.4.1	Local Component Example	70
4.4.2	Global Component Example	71
4.5	Experimental Studies	73
4.5.1	Choosing Proper Parameters	74
4.5.2	Robustness against Collusion Attacks	81
4.5.3	Integration with Dirichlet Reputation System	85
4.5.4	Comparative Experiments	88
4.5.5	Discussions on Experimental Results	93
4.6	Summary	94
5	A Witness Trustworthiness Model	96

5.1	Introduction	96
5.2	Dempster-Shafer Theory Framework	97
5.3	The Witness Trustworthiness Model	100
5.3.1	Personal Trustworthiness	102
5.3.2	Public Trustworthiness	106
5.3.3	Integration	108
5.3.4	A Rating Discounting Scheme	108
5.4	Experimental Studies	110
5.4.1	Experimental Studies for Reputation Systems with Multi-Nominal Rating Levels	111
5.4.1.1	Experimental Settings	111
5.4.1.2	Parameter Value Exploration	113
5.4.1.3	Witness Trustworthiness Investigation	115
5.4.1.4	Robustness Investigation	119
5.4.2	Experimental Studies for Reputation Systems Accepting Only Bi- nary Ratings	120
5.4.2.1	Experimental Settings	121
5.4.2.2	Experimental Results	121
5.5	Discussions	123
5.6	Summary	125
6	Conclusions and Future Work	127
6.1	Contributions	127
6.2	Potential Applications	130

6.2.1	E-Commerce Systems	131
6.2.2	Cognitive Radio Networks	133
6.3	Future Work	133
6.3.1	Improving the Witness Trustworthiness Model	134
6.3.1.1	Elemental Window Size Setting	134
6.3.1.2	Building Incentive Mechanisms to Promote Honesty	135
6.3.1.3	Collecting Testimonies	137
6.3.1.4	Dealing with Subjective Difference	138
6.3.2	Applying to Other Applications	139
6.3.2.1	Review Systems	139
6.3.2.2	Ad-Hoc Networks	140
6.4	Conclusions	141
Author's Publications		143
Bibliography		145

List of Figures

1.1	A typical centralized reputation system	4
1.2	An eBay seller profile	5
2.1	The Beta distribution density function when $\alpha' = 7$ and $\beta' = 1$ [1]	12
2.2	The Bayesian network model for a consumer, two providers and three witnesses, where F_i is a provider feature and θ_f^a denotes the set of variables for all f [2]	22
3.1	(a) MCC value when $T=100$ and $\omega = 100$; (b) MCC value when $T=10$ and $\omega = 100$; (c) FPR value when $T=100$ and $\omega = 100$; (d) FPR value when $T=10$ and $\omega = 100$; (e) FNR value when $T=100$ and $\omega = 100$; (f) FNR value when $T=10$ and $\omega = 100$	50
3.2	(a) MCC value when $T=100$ and $\omega = 10$; (b) MCC value when $T=10$ and $\omega = 10$; (c) FPR value when $T=100$ and $\omega = 10$; (d) FPR value when $T=10$ and $\omega = 10$; (e) FNR value when $T=100$ and $\omega = 10$; (f) FNR value when $T=10$ and $\omega = 10$	51
3.3	(a) U has no transactions; (b) U has transactions	53
3.4	MCC value changes with the number of transactions	55
3.5	(a) P changes his behavior after each 20 transactions; (b) P changes his behavior after each 40 transactions	56

3.6	(a) U has no transactions; (b) U has transactions	58
3.7	(a) $iw = 0.1$; (b) $iw = 0.9$	58
3.8	(a) $iw = 0.1$ and 20% dishonest witnesses; (b) $iw = 0.9$ and 20% dishonest witnesses; (c) $iw = 0.1$ and 90% dishonest witnesses; (d) $iw = 0.9$ and 90% dishonest witnesses	59
4.1	Varying radius when $T=10$, $\omega = 10$, and $P_{unfair} = 40\%$	76
4.2	Varying radius when $T=100$, $\omega = 10$, and $P_{unfair} = 40\%$	77
4.3	Varying radius when $T=10$, $\omega = 100$, and $P_{unfair} = 40\%$	77
4.4	Varying radius when $T=100$, $\omega = 100$, and $P_{unfair} = 40\%$	78
4.5	Varying radius when $P_{unfair} = 80\%$	79
4.6	(a) Frequency of consumer transaction is double as that of witnesses; (b) Equal frequency of transacting with the provider; (c) Frequency of witness transaction is double as that of consumer.	80
4.7	(a) The witnesses' behaviors keep consistent for all providers; (b) The honest witnesses' behaviors keep consistent for all providers; (c) The witnesses' behaviors change over providers; (d) MCC varying with the number of transactions when $P_{unfair} = 60\%$	82
4.8	MAE changes with forgetting factor and the number of windows	87
4.9	MCC changes with the percentage of dishonest witnesses	90
4.10	MCC changes with the number of transactions	90
4.11	Reputation evaluation value changes with the percentage of dishonest witnesses	91
4.12	Reputation evaluation value changes with the number of transactions	92
5.1	Global trustworthiness weight changes with local uncertainty	106

5.2	(a) MAE changes with the number of elemental windows; (b) MAE changes with the percentage of unfair testimonies	114
5.3	Witness trustworthiness	116
5.4	(a) MAE changes with the percentage of dishonest witnesses; (b) MAE changes with the number of the providers that the consumer has no transactions with; (c) MAE changes with the number of the providers that the dishonest witnesses are unfair to	119
5.5	(a) MAE changes with the percentage of unfair testimonies; (b) MAE changes with the percentage of dishonest witnesses; (c) MAE changes with the number of the providers the dishonest witnesses are unfair to	122

List of Tables

2.1	Approach categorization according to the mechanism an approach adopts .	29
2.2	Approach categorization according to the information an approach considers	31
2.3	Approach capabilities	32
3.1	Mapping from willingness value to rating level	48
3.2	Simulation parameters, meanings and values	49
4.1	Rating vectors for the local component working example	71
4.2	The witnesses' honesty regarding respective providers	71
4.3	Relationship between iw value and types of dishonest witnesses	74
4.4	Simulation parameters, meanings and values	75
4.5	Profiles of the providers	83
4.6	Profiles of the providers	89
5.1	Simulation witness types	112
6.1	The mechanism our proposed approaches adopt	131
6.2	The information our proposed approaches consider	131
6.3	The capabilities our proposed approaches have	132

Chapter 1

Introduction

The focus of the research in the field of artificial intelligence has mainly fallen into designing intelligent agents. When these agents co-exist together in a society, Multi-Agent Systems (MAS) are formulated [3]. MAS has emerged as a natural model for large-scale computing systems, such as GRID [4], peer-to-peer computing [5], semantic web [6], web services [7], e-business [8], m-commerce [9], pervasive computing environment [10], autonomic computing [11], massive multi-player online role playing game [12], coordinated defense systems [13], transportation and logistics systems [14], disaster response systems [15], and social network [16].

MAS is an open system, where the agents often exist in an uncertain and dynamic environment. Meanwhile, according to the definition of agent [17] as below:

“a computer system that is situated in some environment, and that is capable of autonomous action in that environment in order to meet its design objectives.”

The independent agents in MAS have the capabilities of autonomous reasoning and decision-making, leading to their self-interest characteristics. The agents interact with each other and may behave deceitfully to achieve their own goals. For example, there may exist agents that do not meet the obligations which are agreed with other agents. Moreover, the agents in MAS can enter and leave the system at any given time, meaning that when an agent does something wrong, it can change its identity then re-enter the

system without being punished. As an example, an agent can sell low-quality product, leave the system as soon as it gets paid, then re-enter the system with a new identity to continue cheating without being known by others. Therefore, there is a risk when an agent engages in transactions with other agents. Even greater risk may be imposed because of the transactions in MAS have some special characteristics which are unseen in the human society. Hence, a mechanism is necessary to regulate the transactions in MAS, and trust is one appropriate choice [18] [19] [20]. Trust is a ubiquitous concept and has many possible meanings. In this thesis, we take one of the most well recognized definitions from [21] as follows:

“a particular level of the subjective probability with which an agent assesses that another agent or another group of agents will perform a particular action, both before he can monitor such action and in a context in which it affects his own actions.”

The research work presented in this thesis falls into the area of modeling trust in MAS. In particular, we focus our attention on addressing trust-related issues for the application domain of online service provision systems. With regard to online service provision systems, a service can represent, for example, an e-commerce transaction (e.g., online seller agents selling items to buyer agents), or a functional component powered by Web Service technologies (e.g., a weather-forecasting component). The research work is motivated by the popularity of online service provision systems. In recent years, with the development of Internet technology, online service provision has prevailed in our daily life. More people use the Internet for entertaining, conducting online transactions, and building personal relationships [22]. As indicated in [23], currently more than 60 hours videos are uploaded by YouTube users who view more than 4 billion videos everyday. It is indicated in [24] that 845 million people all around the world are actively connected with one another on Facebook until February 2012. For the e-business part, a survey [25] has shown that online business has increased from 5.7 billion to over 28 billion in Canada

from 2000 to 2004. In the year 2011, 53% U.S. population bought something online and a total of \$200 billion was spent on online shopping [22]. More people prefer to book their restaurants, hotels and air tickets online. Online service provision has almost touched every part of people’s daily lives and significantly influenced the way of people living [22].

To implement an online service provision system as a multi-agent system, software agents act on behalf of their human users to perform the tasks of conducting transactions. Due to their self-interested characteristics, service provider agents try to maximize their profits when providing service to service consumer agents¹. On the other hand, the consumers try to acquire high quality service at low cost. However, consumers have to rely on some self-enforcing contracts to choose providers to have transactions with due to the lack of complete contracts and legal verification in multi-agent based online service provision systems. Such self-enforcing contracts heavily depend on an evaluation of provider trustworthiness [26]. Therefore, It has become an important topic of research to evaluate provider trustworthiness accurately in online service provision systems [19].

1.1 Motivation

A growing number of research work has been proposed to evaluate the provider trustworthiness in multi-agent based online service provision systems [27] [28] [29] [30] [31]. At an initial stage, only the personal experience information between consumers and providers is taken into account, such as one of the earliest trust model proposed in [32]. Reinforcement learning and game theory are also applied to determine provider trustworthiness by only considering the personal experience information, such as the trust-oriented learning strategy proposed in [33]. However, when a consumer has no personal experience or has only a few personal experiences with providers, the consumer has to depend on other con-

---

¹In the rest of the thesis, we will use “provider” to refer to “service provider agent”, and use “consumer” to refer to “service consumer agent”.

sumers' experiences to evaluate provider trustworthiness. The trust system adopting such kind of strategy is specifically referred to as a *reputation system*, which represents a *soft security* mechanism as a complement to traditional information security [34]. Figure 1.1 shows how a typical centralized reputation system works.

Figure 1.1: A typical centralized reputation system

As Figure 1.1 shows, in an online service provision system, a consumer requests service from a provider and then the provider delivers the service. After the consumer receives the service, he can rate the quality of the service provided and report the rating to the reputation system². The reputation system collects, analyzes and aggregates these reported ratings. Then the provider's reputation score is derived and made available to general consumers. The derived reputation score can be used as an indicator of the trustworthiness of the provider.

The reputation system is widely adopted in online service provision systems. For example, the seller reputation system is adopted in the notable e-commerce system, eBay, to help the buyers (i.e., consumers) to choose reputable sellers (i.e., providers). Figure 1.2 shows an example of a seller's profile in eBay. This seller's reputation is determined based on the ratings from the buyers who had transactions with the seller before. It can

²In a decentralized reputation system, the consumer will report ratings to other consumers when being requested.

be seen that this seller has received 99.1% positive ratings in total. By considering 99.1% as the reputation score and the upper bound of the reputation score is 1, a potential new buyer may perceive a sense that the seller is trustworthy.

Figure 1.2: An eBay seller profile

However, building a reliable reputation system faces many challenging issues. With respect to this, the problem of *unfair testimonies* is one important issue, which happens at the step of consumer reporting ratings to the reputation system in Figure 1.1 and influences the accuracy of provider reputation evaluation as a consequence. For example, suppose there is a consumer U and a provider P . U is now evaluating the reputation of P to decide whether to transact with P using the reputation system which derives P 's reputation score based on the ratings reported by other consumers who had transactions with P before. In particular, we call the consumers who report ratings regarding P as witnesses, and the reported ratings as testimonies³. However, to illude U to transact with P , P might collude with some witnesses who only report positive ratings regarding P no matter what P 's real behavior is. It is highly possible that those unfairly positive ratings will lead to U 's inaccurate evaluation on P 's reputation. As a consequence, U might make

³In the thesis, we use the terms “rating” and “testimony” interchangeably.

a wrong decision to transact with P . Actually, the problem of unfair testimonies exists in every reputation system in the online service provision domains, such as review systems and voting mechanisms. For example, there is one category called “peculiar products” in the “amazon window shop app” for iPad. This category includes many weird and fake products, like uranium ore and a Tuscan whole milk asked for the price of \$40. However, it is surprised that thousands of ratings or reviews are provided for these produces though Amazon does not really sell them. Although most of the reviewers just rate them for fun, it also shows that it is quite easy to manipulate the ratings reported to the reputation system [22]. As another example, in the 2012 university rankings by degree of program, QS rated the program of Communication and Social Studies in Australia National University (ANU) as 24 [35]. But actually ANU does not provide this program.

In particular, unfair testimonies have been adopted by diverse attacks to mislead consumers’ decisions driven by the huge potential profits of online service provision environment in recent years [36]. At the beginning, only the independent unfair ratings are inserted by some individual providers for the purpose of either boosting the providers’ reputation scores or downgrade their competitors’ reputation scores. Then sophisticated programs are developed to make it possible to automatically insert unfair ratings. Nowadays, it is found that there exist some reputation management companies which provide “rating services” for their customers by controlling a large amount of real user IDs. For example, a company named “VideoViralViews.com” [37] can provide 100 real user ratings to a piece of music on iTunes at the cost of about \$750. At the price of \$9.99, a video on YouTube could receive 30 “I like” ratings or 30 real user comments provided by “IncreaseYouTubeViews.com” [22]. More possible ways to mislead consumers’s evaluation on provider reputation are analyzed in [38]. Therefore, without proper defense mechanisms, the existence of unfair testimonies in reputation systems can easily decrease the accuracy of evaluating provider reputation, weaken consumers’ confidence in the reliabil-

ity of reputation systems, and finally undermine the security of online service provision systems [22]. Motivated by the urgent need to build reliable reputation systems, the goal of this thesis is to address the problem of unfair testimonies, which will serve to build reliable reputation systems eventually.

1.2 Contributions

As a summary, the thesis work has made the following major contributions:

1. Though various approaches have been proposed to mitigate the influence of unfair testimonies in reputation systems, most of them are applicable for reputation systems accepting binary ratings only. To address this limitation, we propose a two-stage clustering filtering approach to filter unfair testimonies. Compared with existing approaches, the proposed two-stage clustering filtering approach is applicable not only to the reputation systems supporting binary ratings, but also to those accepting multi-nominal rating levels. Experimental results show that it can outperform the comparative approaches in terms of the accuracy of filtering unfair testimonies and evaluating provider reputation.
2. Furthermore, the two-stage clustering filtering approach is improved as an integrated clustering filtering approach by incorporating more information. The integrated clustering filtering approach considers both the witnesses' testimonies regarding the provider whose reputation is under evaluation, and the witnesses' testimonies regarding other providers. By combining the information together, the integrated clustering filtering approach is applicable to the scenarios where most existing approaches cannot work well, for example, where some witnesses collude together to attack one particular provider (i.e., collusion attacks).
3. Using a mechanism different from the two-stage clustering filtering approach and in-

egrated clustering filtering approach, we further propose a witness trustworthiness model based on Dempster-Shafer theory. The model considers the consumer's personal ratings and the witnesses' testimonies regarding all providers, and then combines different pieces of information together based on Dempster-Shafer theory to calculate the witness trustworthiness value which is specific for different providers. Furthermore, the achieved witness trustworthiness value is used to discount the witness's testimonies to evaluate provider reputation more accurately. Meanwhile, compared with the two-stage clustering filtering approach and integrated clustering filtering approach we propose in the thesis, the witness trustworthiness model has the advantage that it does not heavily depend on an optimized parameter setting to achieve an accurate filtering result. Compared with other existing approaches, the witness trustworthiness model allows the consumer to specify his own subjective difference tolerance level flexibly instead of making an arbitrary decision on which witnesses are trustworthy. Moreover, the witness trustworthiness model can address collusion attacks to a good extent and significantly improve the accuracy of provider reputation evaluation.

1.3 Organization of the Thesis

The thesis is organized as follows.

- In Chapter 2, we review some representative approaches addressing the problem of unfair testimonies. The advantages and disadvantages are also analyzed. Then we categorize the approaches according to the categorization methods we introduce. The capabilities that an approach addressing the problem of unfair testimonies should have are presented, and the factors impacting the design of an approach are analyzed.

- In Chapter 3, we present a two-stage clustering filtering approach to filter unfair testimonies for reputation systems supporting multi-nominal rating levels. We first present how the two-stage clustering filtering approach works. Then we simulate an online service provision environment to explore the accuracy of the two-stage clustering filtering approach in filtering unfair testimonies in various scenarios. We also compare this approach with other approaches by integrating it with a reputation system accepting binary ratings only.
- In Chapter 4, we present an integrated clustering filtering approach to revamp the two-stage clustering filtering approach proposed in Chapter 3. The performance of the integrated clustering filtering approach and comparison with other approaches in terms of the accuracy of filtering and evaluation of provider reputation, are presented through experimental results.
- In Chapter 5, we present a witness trustworthiness model based on Dempster-Shafer theory to address the problem of unfair testimonies by using a mechanism different from the approaches proposed in Chapter 3 and 4. We first introduce how to model a witness's trustworthiness based on Dempster-Shafer theory. Then we present a rating discounting scheme using the witness trustworthiness value calculated from the model. Experimental results are presented to show the effectiveness of the proposed witness trustworthiness model in contributing to building reliable reputation systems.
- In Chapter 6, we first conclude the thesis by highlighting the contributions our work made in addressing the problem of unfair testimonies. Then we discuss the possible future directions.

Chapter 2

Literature Review

In this chapter, a detailed literature survey on the approaches addressing the problem of unfair testimonies to build reliable reputation systems is conducted. The advantages and disadvantages of the approaches are analyzed in details. Then we categorize these approaches from two aspects: the mechanism the approach adopts and the information it uses. The capabilities that the approaches addressing the problem of unfair testimonies should have, and the factors impacting the approach design are also presented.

2.1 Approaches to Address the Problem of Unfair Testimonies

In this section, we present a comprehensive survey on various representative approaches to address the problem of unfair ratings.

2.1.1 Iterated Filtering Approach

Whitby et al. [1] proposed an iterated filtering approach to filter unfair testimonies for Beta Reputation System (BRS) [39], which uses Beta distribution to model provider reputation. In BRS, the provider behavior encountered by each individual consumer is considered as a binary event, which can be modeled by a Beta distribution [40], whose

probability density function is defined as follows:

$$B(p|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}, \quad (2.1)$$

where $\alpha > 0$ and $\beta > 0$ are two parameters determining the shape of a Beta distribution, Γ is the gamma function, and p is a probability variable in the range of $[0, 1]$. This density function shows the relative likelihood of the variable p when the shape parameters α and β are fixed.

In BRS, after a transaction between a consumer and a provider is completed, the consumer can give a positive rating (e.g., expressed as 1) to the provider if the provider behaves cooperatively in the transaction, or a negative rating (e.g., expressed as -1) if the provider does not meet the satisfaction, and report the rating to BRS. Then BRS is able to estimate the reputation of the provider by simply accumulating the propagated ratings provided by different consumers. Suppose the accumulated number of positive ratings and negative ratings is α' and β' , respectively. By using Laplace-estimator [41], BRS sets the two shape parameters in Beta distribution function as [39] [42]:

$$\alpha = \alpha' + 1, \quad (2.2)$$

$$\beta = \beta' + 1, \quad (2.3)$$

Then the provider's reputation E is estimated as the probability expectation value of p as follows:

$$E = \frac{\alpha}{\alpha + \beta}. \quad (2.4)$$

Figure 2.1 shows an example of the Beta probability density function when $\alpha' = 7$ and $\beta' = 1$ [1]. According to Equation (2.4), the provider's reputation who has the Beta distribution density function as shown in Figure 2.1 is 0.8.

To address the problem of unfair testimonies in BRS, the iterated filtering approach

Figure 2.1: The Beta distribution density function when $\alpha' = 7$ and $\beta' = 1$ [1]

works as follows. It first calculates the provider reputation using the ratings from a particular witness only. Then it calculates the provider reputation using the ratings provided by the majority witnesses, which are initially all witnesses. If the first value is outside of the q quantile or $1 - q$ quantile of the second value, the particular witness is identified as dishonest and excluded from the set of majority witnesses. The filtering process continues until the set of the majority witnesses does not change any more. Figure 2.1 illustrates the major point of the filtering process when q is set as 0.01. Suppose the Beta distribution density function has the shape as shown in Figure 2.1 according to majority witnesses' ratings. The calculated provider reputation according to one particular witness's ratings is E' . If the integral of p from negative infinity to E' is smaller than 0.01 or greater than 0.99 (i.e., $E' < 0.456$ or $E' > 0.983$ in Figure 2.1), the witness is considered as dishonest and excluded from the majority witness set.

Though the iterated filtering approach is very simple, it has the disadvantages of time-consuming and especially, being sensitive to the percentage of unfair testimonies, which makes it perform poorly when there exist many unfair witnesses. For example, suppose there are four witnesses, w_1 , w_2 , w_3 and w_4 . Each of them provides one rating. w_1 and w_2 are honest witnesses and each provides a positive rating. w_3 and w_4 are dishonest

witnesses and each provides a negative rating. By setting q as 0.01, all the four witnesses are classified as honest witnesses after the iterated filtering process.

2.1.2 Entropy-Based Filtering Approach

Weng et al. proposed an entropy-based approach [43] to filter unfair testimonies for BRS. This approach first calculates the quality of a testimony R according to the information uncertainty the testimony contains. Then it compares the quality of the testimony R with that of identified honest witnesses' ratings to filter unfair testimonies.

More specifically, the information uncertainty is measured by entropy [44]. The uncertainty contained in a testimony R is given by the following equation:

$$H(R) = -Pr(\alpha)\log(Pr(\alpha)) - Pr(\beta)\log(Pr(\beta)), \quad (2.5)$$

where $Pr(\alpha)$ and $Pr(\beta)$ are given by:

$$Pr(\alpha) = \frac{\alpha}{\alpha + \beta}, \quad (2.6)$$

$$Pr(\beta) = \frac{\beta}{\alpha + \beta}, \quad (2.7)$$

where α and β is the number of positive ratings and negative ratings contained in the testimony, respectively. Then the quality of the testimony R is given by the following equation:

$$Q(R) = 1 - \frac{H(R) - H_{min}}{H_{max} - H_{min}} = \frac{H_{max} - H(R)}{H_{max} - H_{min}}, \quad (2.8)$$

where H_{max} is achieved from Equation (2.5) when the numbers of positive ratings and negative ratings are supposed to be equal to each other, and H_{min} is achieved from Equation (2.5) by supposing that there are only positive ratings or negative ratings in the testimony.

Suppose the personal ratings from the consumer himself is R_U and the quality for

the personal ratings is also measured using Equation (2.8) as $Q(R_U)$. The entropy-based filtering approach calculates the absolute difference between $Q(R)$ and $Q(R_U)$. If the absolute difference value exceeds a threshold ε , then the testimony R is considered as unfair and discarded. Otherwise, R is accumulated with the consumer's personal ratings. In the following filtering process, the quality of the accumulated ratings rather than the consumer's personal ratings are used to compare with the quality of another testimony.

This approach has the advantage that it proposes the concept of testimony quality by using entropy. But it has some problems. First, it assumes that the consumer has personal ratings. However, it is highly possible that the consumer needs to evaluate the provider's reputation when his personal ratings are lacking. Second, because of using entropy, the approach cannot distinguish the quality between the symmetry positive and negative rating pairs, which may lead to that unfair testimonies cannot be accurately identified. For example, suppose a testimony contains 2 positive ratings and 8 negative ratings. The consumer himself has 8 negative ratings and 2 positive ratings. The two rating pairs have the same quality values according to Equation (2.5) and (2.8) though they are obviously two different rating pairs.

2.1.3 Clustering Filtering Approach

Dellarocas proposed a clustering filtering approach [45] to address the problem of unfair testimonies for online trading system. In this approach, the unfair testimonies are separated into two groups – unfairly low ratings and unfairly high ratings. The unfairly low ratings are avoided by the mechanism of controlled anonymity. To mitigate the influence of unfairly high ratings, collaborative filtering techniques [46] are first employed to identify the consumer's nearest neighbor, which includes other consumers whose ratings are similar to the consumer's ratings for the commonly rated providers. Second, a divisive clustering technique is used to further filter unfairly high ratings by separating the neigh-

bor's ratings into two clusters – the cluster including the lower ratings and the cluster including the higher ratings. Then the ratings in the cluster including higher ratings are considered as unfair and discarded as unfairly low ratings are assumed to be avoided.

The clustering filtering approach has the advantage that it only considers the ratings provided within the most recent dynamic time window, whose width is adapted to the frequency of fair ratings reporting. Therefore, it can cope with the situation where providers vary behaviors over time. But this approach has the disadvantage that it actually cannot address unfairly low ratings effectively. It is pointed out in this work that unfairly high ratings cannot be addressed by the mechanism of controlled anonymity as consumers and providers are always able to find a way to signal their identities to each other if they are willing to do so even when their identities are concealed. For example, a provider may use a specific pattern in the amounts of his services in order to signal his presence to some consumers who collude with him. Nevertheless, since consumers can provide ratings for all providers, it is highly possible that controlled anonymity by concealing identity signal cannot control unfairly low ratings.

2.1.4 Two-Layer Cognitive Filtering Approach

Noorian et al. proposed a two-layer filtering approach that cognitively elicits the behavioral characteristics of witnesses [47]. As the approach name suggests, it has two layers. In the first layer, the consumer U measures the competency of his neighbors to identify potential honest witnesses. The consumer first sends a query to his neighbors requesting the testimonies for the providers the consumer has encountered before. Suppose the testimony R_{N_k, P_j} for a provider P_j from a witness N_k is a tuple of $\langle r, s \rangle$ which indicates the numbers of positive ratings (i.e., r) and negative ratings (i.e., s). P_j 's reputation $E(P_j)_{N_k}$ using the testimony from N_k only is evaluated as $\frac{r+1}{r+s+2}$ by using BRS. Then the average

level of dishonesty of N_k is calculated by using the following equation:

$$\bar{d}_{N_k} = \frac{\sum_{j=1}^{|P|} |E(P_j)_U - E(P_j)_{N_k}|}{|P|}, \quad (2.9)$$

where P is the number of the providers U requests testimonies for and $E(P_j)_U$ is P_j 's evaluated reputation by using U 's personal ratings only.

To address the situation where an honest witness lacks experiences, and therefore, is not eligible to play the role of witness, the reliability degree of each R_{N_k, P_j} is defined as follows [41]:

$$c(r, s)_{N_k, P_j} = \frac{1}{2} \int_0^1 \left| \frac{x^r (1-x)^s}{\int_0^1 x^r (1-x)^s dx} - 1 \right| dx. \quad (2.10)$$

To measure the information imbalance between U and N_k 's ratings, an uncertainty function $\bar{U}_{N_k}$ is further defined as:

$$\bar{U}_{N_k} = \frac{\sum_{j=1}^{|P|} |c(r, s)_{U, P_j} - c(r, s)_{N_k, P_j}|}{|P|}. \quad (2.11)$$

Given Equation (2.9) and (2.11), the witness N_k 's competency degree $Comp_{N_k}$ is calculated through his honesty based on his uncertainty degree as follows:

$$Comp_{N_k} = (1 - \bar{d}_{N_k})(1 - \bar{U}_{N_k}). \quad (2.12)$$

U will only choose the witnesses whose competence degree is lower than a predefined incompetency tolerance threshold μ as potential honest witnesses and pass them to the second filtering layer.

In the second layer, the consumer U calculates the average similarity between the criteria preferences of his personal ratings and those of the witness N_k 's ratings. Denote the similarity for a provider P_j be Sim_{U, N_k, P_j} . Then the difference for the criteria preferences $Diff_{U, N_k, P_j}$ is calculated as $1 - Sim_{U, N_k, P_j}$. The average differences $\overline{Diff}_{U, N_k}$ between

the criteria preferences of U and N_k is calculated as follows:

$$\overline{Diff}_{U,N_k} = \frac{\sum_{j=1}^{|P|} |Diff_{U,N_k,P_j}| \times T_{U,N_k,P_j}}{|P|}, \quad (2.13)$$

where T_{U,N_k,P_j} is a time decaying factor [39] achieved from the distance between the latest ratings of U and N_k regarding P_j . Furthermore, to embrace the diversity of the witnesses' behaviors in providing ratings, U captures the overall tendency $Tendency_{U,N_k}$ of N_k in providing ratings as:

$$Tendency_{U,N_k} = \frac{\sum_{j=1}^{|P|} Diff_{U,N_k,P_j}}{|P|}. \quad (2.14)$$

By considering both the average difference and tendency results, the witnesses' behavior patterns are classified as credible, optimistic, pessimistic and deceitful. Given the witnesses' behavior patterns and average difference results, the witnesses' credibility measures are finally achieved.

This approach has the advantage that it takes into account the diversity of the witnesses' behaviors and classifies the witnesses accordingly. However, it cannot work when the witnesses' behaviors vary over providers. For instance, a witness may provide fair ratings for some providers, and provide unfair ratings for some other providers. As the approach calculates the witness's average competency degree for all providers, it cannot tell to which providers the witness are dishonest. The witness may also behave optimistically for some providers and behave pessimistically for some other providers. The average tendency and difference values cannot tell what specific behavior pattern the witness presents for a particular provider.

2.1.5 TRAVOS

TRAVOS [48] is a reputation model based on the Beta probability density function. It copes with the problem of unfair testimonies by accomplishing two tasks. First, it es-

timates the accuracy of a witness's testimonies by comparing the witness's ratings with the consumer's personal ratings regarding all providers. In particular, TRAVOS divides the interval of $[0,1]$ into N disjoint bins. It then uses the witness's current testimony regarding the target provider (i.e., the provider whose reputation is under the consumer's evaluation) to calculate the provider's reputation using BRS and locate the bin that the reputation value falls into, denoted as bin^o . Then TRAVOS calculates all providers' reputation by applying BRS using the testimonies from the witness regarding all providers. If one provider's reputation value falls into bin^o , TRAVOS will check whether the consumer's rating regarding this provider is positive or negative. TRAVOS aggregates the number of positive ratings and negative ratings from the consumer as C_{pos} and C_{neg} , respectively. Then the accuracy of the witness's testimonies ρ is calculated as follows:

$$\rho = \frac{\int_{\min(bin^o)}^{\max(bin^o)} X^{\alpha^o-1} (1-X)^{\beta^o-1} dX}{\int_0^1 U^{\alpha^o-1} (1-U)^{\beta^o-1} dU}, \quad (2.15)$$

where $\alpha^o = C_{pos} + 1$ and $\beta^o = C_{neg} + 1$.

Consider the following example to estimate the accuracy of a witness's testimonies. Suppose the interval $[0,1]$ is divided into 5 bins which are $[0, 0.2]$, $[0.2, 0.4]$, $[0.4, 0.6]$, $[0.6, 0.8]$ and $[0.8, 1]$. Suppose the witness's current testimony regarding a provider includes 7 positive ratings and 1 negative ratings. Using BRS, the provider's reputation value falls into bin^5 . Suppose the consumer has 3 positive ratings and 0 negative rating after scanning the witnesses' testimonies for all providers. The accuracy of the witness's testimonies is then calculated as 0.5904 by substituting $\alpha^o = 4$ and $\beta^o = 1$ into Equation (2.15).

To mitigate the influence of the attack that the dishonest witnesses may flood the reputation system by providing a large amount of positive or negative ratings, TRAVOS performs the second task to adjust the witnesses' testimonies according to the accuracy achieved in the first task. Then the adjusted testimonies will be passed to BRS to evaluate

the provider's reputation.

Though TRAVOS achieves better performance than the iterated filtering approach as the experimental results presented in [48], it has some problems. First, sharing the similar disadvantage as the entropy-based filtering approach, it requires the consumer's personal ratings to calculate the accuracy of the witnesses' testimonies. However, the personal ratings are unavailable for a consumer who is a newcomer to the reputation system, resulting in that TRAVOS cannot work in such situation. Second, it assumes that providers behave consistently, but this assumption may not be true. It is highly possible that a provider may change his behavior over time. Suppose an honest witness who had transaction with a provider before and provided a positive rating for this provider. However, this rating may be incorrectly considered as unfair if the consumer had transaction with the provider after the provider changed his behavior as TRAVOS works based on comparing the consumer's personal ratings with the witness's ratings. The third problem is that TRAVOS is quite time-consuming as it will scan all witnesses' testimonies regarding all providers everytime when the witnesses' accuracy is estimated.

2.1.6 Yu and Singh's Approach

Yu and Singh proposed a model [49] to evaluate the trustworthiness of a provider based on Dempster-Shafer theory [50] [51]. The model focuses on the following two jobs:

- Collecting witness testimonies.
- Combining collected witness testimonies together.

The first job is completed by using referral networks [52]. Each consumer maintains a set of acquaintances, and some of the acquaintances are identified as his neighbors. Each consumer also maintains a model of each acquaintance, which describes the acquaintance's

trustworthiness and helps the consumer to trace other trustworthy consumers through referrals. Therefore, the function of the referral network is to facilitate the consumer to find reliable witnesses regarding a particular provider by tracing referrals from his neighbors.

The second job is completed by using Dempster-Shafer theory. More specifically, each consumer defines a lower bound ω and an upper bound Ω . Then the distribution of a witness's testimony is divided into three regions by using ω and Ω :

- The region that contributes to the consumer's evaluation that the provider is untrustworthy.
- The region that contributes to the consumer's evaluation that it is uncertain regarding whether the provider is trustworthy or untrustworthy.
- The region that contributes to the consumer's evaluation that the provider is trustworthy.

The ratings in different regions are counted to derive a Basic Belief Assignment (BBA) [51] over the distribution of the ratings. The provider's trustworthiness is evaluated by combining the BBAs from the consumers using Dempster-Shafer combination rule [51].

The presence of unfair testimonies is addressed by introducing a variant of the weighted majority algorithm (WMA) [53] – continuous WMA (WMC) [54]. With WMC, each consumer maintains a weight for each witness whose testimonies have been aggregated. The weight can be considered as the trustworthiness of the corresponding witness. Initially, each witness's weight is assigned as 1. Then each witness's weight is tuned by comparing the provider trustworthiness evaluation value using the witness's ratings only with that using the consumer's personal ratings only. The witness's testimony is further discounted using his weight. The provider trustworthiness is finally evaluated by combining

the discounted testimonies using Dempster-Shafer combination rule.

Although this work does not depend on extra mechanisms or knowledge, it suffers from some disadvantages. First, each witness's weight is calculated through comparing the consumer's personal ratings with the witness's ratings for the provider only. Therefore, the weight of the witness will not be tuned if the consumer has no much personal experience with the provider. The consumer will take the witness's testimonies into account heavily. Second, WMC is very sensitive to unfair testimonies. If a testimony is not exactly the same as the consumer's personal ratings, the witness' weight will be decreased and the testimony will be discounted accordingly. As a result, in the scenario where most testimonies are reliable, the fair testimonies' contribution to the provider trustworthiness evaluation is reduced and the model works even worse than not applying WMC. Another problem of this approach is that once a witness's weight is decreased, there is no mechanism to make the witness gain his weight back from the consumer.

2.1.7 Bayesian Learning Approach

Regan et al. proposed the BLADE model [2] to cope with the problem of unfair testimonies by using Bayesian learning technique [55]. The BLADE model tries to reinterpret a witness's ratings instead of filtering or adjusting his ratings by assuming that the witness biases his ratings consistently. More specifically, the consumer can establish a Bayesian network using his personal ratings and the witness's testimonies. For example, Figure 2.2 illustrates a Bayesian network constructed by the consumer when there are two providers (i.e., s_1 and s_2) and three witnesses (i.e., a_1 , a_2 and a_3). Each provider has three features to be evaluated (e.g., in an e-marketplace environment, the features can be item quality, item price and item delivering time).

The task of learning is to find the witness's evaluation function θ_f^a when the provider exhibits the feature values f . Initially, the consumer starts with the priors $Pr(\theta_i^s)$ and

Figure 2.2: The Bayesian network model for a consumer, two providers and three witnesses, where F_i is a provider feature and θ_f^a denotes the set of variables for all f [2]

$Pr(\theta_f^a)$, which encodes the likelihood of the behavior for each provider and the evaluation function for each witness, respectively. As the consumer has more transactions with the providers and receives more testimonies from the witnesses, the observed provider features f^s are used to calculate the posteriors of $Pr(\theta_i^s|f^s)$ and $Pr(\theta_f^a|f^s)$. The reported testimonies r_s^a are used as evidence to calculate the posteriors $Pr(\theta_i^s|r_s^a)$ and $Pr(\theta_f^a|r_s^a)$.

For example, suppose a consumer has 10 transactions with a provider s_1 . Among the 10 transactions, the consumer observes that the provider delivered the items *on time* 9 times and *late* once. The consumer also collects ratings from a witness regarding s_1 . Suppose the witness provides 1 positive rating and 9 negative ratings regarding s_1 . By using BLADE, the consumer infers that there is a correlation between the witness's negative rating and the provider's feature value *on time*. Then suppose the consumer is evaluating the reputation of another provider s_2 with whom the consumer totally has no transactions. And the witness provides a testimony regarding s_2 containing 9 positive

ratings and 1 negative rating. Because the consumer already learned that the witness's evaluation function, the consumer is able to reinterpret the witness's positive rating as s_2 delivering items late and negative rating as s_2 delivering items on time.

As an extension, Teacy et al. proposed the HABIT model to extend BLADE model to a general application area with an appropriate parameter setting [56]. But both BLADE and HABIT have some inherent disadvantages. First, BLADE requires that the consumer has personal ratings with the providers already. Therefore, it cannot work when a newcomer enters the system. Second, it infers the witness's evaluation function using probability distribution. Suppose a witness reports 5 negative ratings which happen when the provider delivers items on time, and 5 positive ratings which happen when the provider delivers items late. BLADE will fail to get the correlation between negative rating and *on time*, or positive rating and *late*. Furthermore, BLADE assumes that the witnesses behave consistently over all providers, which makes it not applicable to the scenarios where witness behaviors vary over providers.

2.1.8 CHI-Score Approach

In [57], Jura and Faltings proposed a trust model by using Chi-square independence test [58]. The problem of unfair testimonies is addressed by rewarding witnesses to elicit fair ratings. More specifically, this approach first uses Chi-square independence test to establish whether a witness's testimony is different from other witnesses' testimonies regarding the target provider. Then instead of outputting a binary decision on whether the witness is trustworthy or not, a continuous value (called CHI-score) is output by modifying the Chi-square independence test. The witness's CHI-score is used in two ways. First, it is used as an indicator to identify unfair testimonies. The higher the CHI-score is, the testimonies under test are more likely to be fair. Second, it is used as a side-payment weight for the witness providing testimonies to encourage the witnesses to report ratings

in an honest way.

This approach has the advantage that it can be easily extended to reputation systems supporting multi-nominal ratings [59]. But due to using Chi-square independence test, it may require that the witness has sufficient number of ratings to ensure the statistical accuracy of the test. Moreover, it compares the witness's ratings with other witnesses' ratings in the Chi-square independence test, resulting in that it will fail when majority witnesses are dishonest. In addition, a side-payment mechanism is required to be implemented, which is separated from the ordinary payment mechanism used in consumers and providers' transactions, and hence, external overhead is required.

2.1.9 Personalized Approach

Zhang and Robin proposed a personalized approach to address the problem of unfair testimonies [60] [61]. This approach takes into account a consumer's personal ratings and other witnesses' testimonies to model a witness's reputation when providing testimonies. More specifically, this approach first calculates a witness's private reputation by comparing the consumer's personal ratings with the witness's testimonies regarding their commonly rated providers. Suppose the number of the similar ratings and different ratings between the consumer and the witness regarding the commonly rated providers is α and β , respectively. Then the witness's private reputation is calculated as $\frac{\alpha+1}{\alpha+\beta+2}$.

Second, the approach calculates the witness's public reputation based on the common and centrally held ratings regarding the providers in the system. In particular, the witness's public reputation is calculated by comparing the witness's ratings with other witnesses' ratings for the providers the witness encountered before. Suppose the number of the similar ratings and different ratings between the witness and other majority witnesses is α' and β' , respectively. Then the witness's public reputation is calculated as $\frac{\alpha'+1}{\alpha'+\beta'+2}$.

Finally, the witness's reputation is obtained by combining the calculated private and public reputation together. Suppose the maximal level of error the consumer can accept is ε and the confidence measure is γ , then according to Chernoff Bound theorem [62], the minimum number of ratings N_{min} that the consumer and the witness share for the commonly rated providers is determined as follows:

$$N_{min} = -\frac{1}{2\varepsilon^2} \ln \frac{1-\gamma}{2}. \quad (2.16)$$

The weight ω for the private reputation is assigned as:

$$\omega = \min\left(\frac{N_{all}}{N_{min}}, 1\right), \quad (2.17)$$

where N_{all} is the number of the ratings in the private reputation part that the consumer and the witness share for the commonly rated providers. The weight for the witness's public reputation is assigned as $1 - \omega$. The witness's reputation is calculated as the weighted sum of the private and public reputation. Furthermore, according to subjective logic [63], the witness's reputation is used to adjust his testimonies by using a rating discounting scheme and a revised version of BRS is developed to evaluate the provider's reputation [61].

This approach has the advantage that it considers both the witness's private and public reputation, which makes it applicable in some scenarios where other approaches cannot work, for example, when majority witnesses provide unfair ratings. However, it has some disadvantages. First, it is designed for reputation systems accepting binary ratings only, leading to that it cannot work for the reputation systems supporting multi-nominal rating levels [59]. Second, similar to one disadvantage of TRAVOS and BLADE, it models the witness's reputation as a general value for all providers. Therefore, it cannot address the scenario where the witnesses change their behaviors over providers.

2.1.10 Trust-Based Incentive Approach

Zhang and Cohen proposed a trust-based incentive approach to promote honesty [64]. The honesty includes the provider's honesty as well as the consumer's honesty in reporting ratings. In addition, it tries to encourage the consumers to provide ratings to address the dilemma that the consumers have no direct incentives to provide ratings [65] [66]. The consumer's honesty is modeled by counting the number of other consumers considering him as a trustworthy witness. More specifically, a central server is employed to maintain a fixed number of neighbors for each consumer, which are the most reputable witnesses the consumer considers when providing testimonies. A witness's reputation is achieved through the personalized approach we reviewed in Section 2.1.9. The central server also provides the neighbor list information to the providers when being requested. Then the providers count the number of a consumer being another consumer's neighbor. Suppose that a consumer B is considered as neighbors by other N_B consumers, the consumer's honesty $R(B)$ is calculated as follows:

$$R(B) = \min\left(\frac{N_B}{\theta}, 1\right), \quad (2.18)$$

where θ is the number of consumers in the system.

When a consumer broadcasts that he needs a service, a provider will bid for the service by taking into account the consumer's honesty and the provider's expected profit from winning this auction. The bids provided by the providers will be submitted to the central server and passed to the consumer. The consumer first calculates the providers' reputation using the approach proposed in [61]. Only the providers whose reputation values exceed a threshold are selected as potential partners. Then the consumer calculates the expected profits by considering the bids providing by the potential partners. The provider who is reputable and whose bid produces the largest profit is selected as the winner.

This approach has the advantage that it not only models the witness's reputation and the provider's reputation but also models the consumer's honesty. But because it requires a central server to coordinate all tasks, the central server may become a single-point failure. Furthermore, if there are a large amount of consumers in the system, the consumers' neighbor lists will hardly overlap with each other, resulting in that each consumer's honesty is quite low. In addition, it inherits the disadvantages of the personalized approach as it uses the personalized approach to model the witness reputation.

2.2 Categorization of the Approaches

In the previous section, we have reviewed various representative approaches proposed to address the problem of unfair testimonies. In this section, we present a categorization of these approaches from two aspects – the mechanism an approach adopts and the information an approach takes into account.

From the aspect of the mechanism an approach adopts, there are four dimensions – filtering, modeling, learning and incentive mechanisms.

- **Filtering mechanism:** a filtering mechanism approach tries to isolate the unfair ratings from the fair ratings using some threshold values and discard the unfair ratings accordingly. For example, the iterated filtering approach will discard the ratings which are not within $q \sim 1 - q$ quantile of the majority ratings. The entropy-based filtering approach will discard the ratings whose quality is different from the quality of the aggregated testimonies at more than the predefined threshold value. In the two-layer cognitive filtering approach, a witness's competence is first calculated to filter the most dishonest witnesses.
- **Modeling mechanism:** a modeling mechanism approach tries to model witness trustworthiness. Then the witnesses' trustworthiness values can be used to adjust the

witnesses' testimonies, reward the witnesses to encourage honest reporting, or filter the witnesses according to some threshold values. For example, in the TRAVOS and personalized approach, a witness's trustworthiness is first estimated. Then the witness's trustworthiness value is used to adjust his reported ratings. In the CHI-score approach, a witness's CHI-score is first calculated as his trustworthiness and further used to reward the witness accordingly.

- **Learning mechanism:** a learning mechanism approach tries to learn the behavior pattern of a witness and then reinterpret his reported ratings according to the learned behavior. For example, in the Bayesian learning approach, a witness's rating pattern is first learned, such as that the witness always reports negative ratings when a provider actually behaves well and reports positive ratings when the provider behaves bad. Then another provider's behavior can be inferred by reinterpreting the witness's reported ratings. For instance, if the witness reports 10 negative ratings regarding another provider, then the provider's behavior can be implied as behaving well by reinterpreting the 10 negative ratings as that the provider behaves well 10 times according to the learned pattern regarding the witness reporting ratings.
- **Incentive mechanism:** an incentive mechanism tries to reward witnesses to encourage the witnesses to report testimonies honestly. For example, in the trust-based incentive approach, more honest a consumer is in reporting ratings, more discounts can be obtained from providers and more profits can be achieved as a consequence.

It is worth pointing out that the approaches from different dimensions can be compatible. For example, a modeling approach can also be a filtering approach if it discards the testimonies based on the predefined threshold values. A modeling approach can also be an incentive approach if it adopts some strategy to reward the witness according to the calculated witness trustworthiness value. Table 2.1 shows the categorization of the

reviewed approaches to cope with the problem of unfair testimonies considering from the aspect of the mechanism an approach adopts. In this table, the symbol $\checkmark$ represents that the approach adopts the mechanism when it is originally designed. The symbol $\approx$ represents that the approach is also compatible to the mechanism after some revising.

Approach	Filtering	Modeling	Learning	Incentive
Iterated Filtering	$\checkmark$			
Entropy-Based Filtering	$\checkmark$			
Cluster Filtering	$\checkmark$			
Multi-Layer Cognitive Filtering	$\checkmark$	$\approx$		
TRAVOS	$\approx$	$\checkmark$		
Yu and Singh	$\approx$	$\checkmark$		
Bayesian Learning			$\checkmark$	
CHI-Score	$\approx$			$\checkmark$
Personalized	$\approx$	$\checkmark$		
Trust-Based Incentive	$\approx$	$\approx$		$\checkmark$

Table 2.1: Approach categorization according to the mechanism an approach adopts

From the aspect of the information an approach takes into account, there are the following four dimensions according to the categorization proposed in [60]:

- **Personal:** an approach is personal if it addresses the problem of unfair testimonies based on comparing the consumer's personal ratings with the witness's past testimonies. It is highly possible that a witness's current rating is also fair if his past testimonies are fair. For example, in the TRAVOS model, the accuracy of the witness's current rating is estimated by comparing the witness's past ratings with the consumer's personal ratings. An approach considering only personal information has the disadvantage that it cannot work when the consumer's personal ratings are missing.
- **Public:** an approach is public if the approach uses all witnesses' testimonies regarding the target provider to cope with the problem of unfair testimonies. A witness's testimonies are likely to be fair if his testimonies are similar to majority witnesses'

ratings for the same provider. For example, in the iterated filtering approach, the testimonies that are not within $q \sim 1 - q$ quantile of the majority testimonies are filtered out as unfair testimonies. An approach considering only public information has the problem that it will perform poorly when majority witnesses are dishonest.

- Local: an approach is local if it uses the ratings regarding the target provider to address the problem of unfair testimonies. For example, the iterated filtering approach and the clustering filtering approach filter the unfair ratings regarding the target provider based only on the ratings regarding this provider. An approach considering only local information also has the problem that it cannot work when majority witnesses are dishonest.
- Global: an approach is global if it uses the ratings for all the providers that a witness has ever rated. For example, TRAVOS is a global approach as a witness's trustworthiness is estimated by comparing the witness's ratings with the consumer's personal ratings regarding all providers. An approach considering only global information will have the problem that it cannot address the scenario where witness behaviors vary over providers.

The categorization of the reviewed approaches to address the problem of unfair testimonies is summarized in Table 2.2 from the aspect of the information an approach uses. It can be noticed that no approach makes use of the four dimensions of information together.

2.3 Capabilities

In Section 2.1 and 2.2, we have reviewed various representative approaches handling the problem of unfair testimonies, analyzed the advantages and disadvantages these approaches have, and categorized them from two categorization aspects. In this section, we

Approach	Personal	Public	Local	Global
Iterated Filtering		✓	✓	
Entropy-Based Filtering	✓	✓	✓	
Cluster Filtering		✓	✓	
Multi-Layer Cognitive Filtering	✓			✓
TRAVOS	✓			✓
Yu and Singh	✓		✓	
Bayesian Learning	✓			✓
CHI-Score		✓	✓	
Personalized	✓	✓		✓
Trust-Based Incentive	✓	✓		✓

Table 2.2: Approach categorization according to the information an approach considers

will generally discuss the capabilities an effective approach should have. By extending the capabilities proposed in [60], we list the following five capabilities:

- **Majority:** an effective approach should be able to address the scenario where majority ratings of a provider are unfair. For example, the performance of the iterated filtering approach decreases quickly with the increasing of the percentage of unfair ratings. On the contrary, TRAVOS can work when majority witnesses are unfair.
- **Flooding:** an effective approach should have the capability to address the scenario where unfair ratings arrive within a short time frame. For example, the iterated filtering approach may perform poorly in this kind of situation, but TRAVOS has the ability to deal with this scenario.
- **Lack:** an effective approach should have the capability to address the scenario where the consumer has no or only a few personal transactions with providers. For example, TRAVOS cannot work when the consumer is a newcomer to the system. But the iterated filtering approach can work when the consumer's personal ratings are unavailable.
- **Provider Behavior Varying:** the performance of an effective approach should not be

influenced by provider behavior changes. Witnesses may provide different ratings for the same providers when provider behaviors vary. The different ratings due to provider behavior changes should not be classified as unfair ratings. For example, though the iterated filtering approach itself does not have this capability, it leaves this problem to BRS by implying a time forgetting factor [39]. But TRAVOS cannot deal with provider behavior varying.

- **Witness Behavior Varying:** an effective approach should be able to deal with witness behavior changes over providers. Witnesses may report fair ratings for some providers, and report unfair ratings for other providers. An even more serious situation is that some witnesses strategically collude together to attack some particular providers. The iterated filtering approach has this capability as it uses the ratings regarding the target provider only to filter unfair testimonies. But TRAVOS will be influenced by witness behavior changes as it estimates the trustworthiness of a witness by generally considering the witness's ratings regarding all providers.

Table 2.3 lists the capabilities of the approaches reviewed in Section 2.1.

Approach	Majority	Flooding	Lack	Provider Varying	Witness Varying
Iterated Filtering			✓	✓	✓
Entropy-Based Filtering	✓	✓		✓	✓
Clustering Filtering		✓	✓		✓
Two-Layer Cognitive Filtering	✓	✓			
TRAVOS	✓	✓			
Yu and Singh	✓	✓			✓
Bayesian Learning	✓	✓			
CHI-Score		✓	✓		✓
Personalized	✓	✓	✓	✓	
Trust-Based Incentive	✓	✓	✓	✓	

Table 2.3: Approach capabilities

2.4 Factors Impacting the Approaches

In this section, we will discuss some factors that impact the design of the approaches addressing the problem of unfair ratings. The factors impacting the design of the approaches coping with the problem of unfair testimonies are related to the factors for the design of reputation systems as the ultimate goal of the approaches addressing the problem of unfair testimonies is to build reliable reputation systems. Therefore, in this section, we will discuss the impacting factors in the context of designing reputation systems.

2.4.1 System Architecture

The first factor involved in building a reputation system is system architecture – whether the system is centralized or decentralized [20]. When a centralized architecture is adopted, there is a central entity in the system to collect and store all the ratings between consumers and providers. The central entity also takes the responsibility of evaluating the reputation of the providers in the system based on the stored ratings. When a consumer needs to evaluate a provider's reputation, he will submit a request to the central entity. Then the central entity will evaluate the reputation of the corresponding requested provider based on the stored ratings, and report the reputation back to the consumer. After a transaction between a consumer and a provider is completed, the consumer will update his rating for this transaction to the central entity. Therefore, all the reputation related information about the providers is shared in the system. When different consumers request the reputation for the same provider, they will get the same result. A typical example of the centralized reputation system is the one adopted in eBay. BRS is also a centralized reputation system.

When the decentralized architecture is adopted, each individual consumer will maintain his past ratings regarding the providers in his own memory. Each individual consumer

will also take the responsibility of computing the reputation of providers by himself rather than depending on a central entity. As each consumer calculates the reputation for the same provider based on his own available information, the reputation results may be different. For example, Yu and Singh's approach is a decentralized reputation system, where each consumer maintains a referral network to collect testimonies and evaluate provider reputation.

2.4.2 Rating Level Granularity

The second impacting factor is the information granularity a reputation system supports [20]. The information granularity refers to the granularity of the feedbacks for providers provided by consumers in transactions. The feedbacks that most reputation systems use are in the form of numeric ratings, which can be discrete or continuous. Discrete ratings can be binary or multi-nominal. In the reputation systems that use discrete ratings, the feedbacks for the providers can be expressed in the form of an integer value, for example, when binary ratings are adopted, 1 represents that the provider behaves well and -1 represents that the provider behaves bad. A continuous rating is a value drawn from a bounded range (usually $[0, 1]$). Continuous ratings can be considered as the finest granularity of discrete ratings. Among the approaches we have reviewed, the iterated filtering approach, entropy-based filtering approach, two-layer cognitive filtering approach, TRAVOS, Bayesian learning approach, personalized approach, and trust-based incentive approach only support reputation systems adopting binary ratings. The CHI-score approach can be extended to support reputation systems accepting multi-nominal ratings. Yu and Sing's approach supports the reputation systems adopting continuous rating levels.

2.5 Summary

In this chapter, we reviewed some representative approaches addressing the problem of unfair ratings. As we have reviewed, though various approaches have been proposed, they suffer from some limitations.

First, most approaches, such as the iterated filtering approach, entropy-based filtering approach, two-layer cognitive filtering approach, TRAVOS, Bayesian learning approach, personalized approach, and trust-based incentive approach are only applicable for the reputation systems accepting binary ratings. However, a reputation system using binary ratings excludes the possibility of providing ratings in multi-levels [59]. Therefore, reputation systems using multi-nominal ratings are proposed to allow consumers to rate providers in a multi-nominal way [59]. But the approaches specifically designed for the reputation systems accepting only binary ratings cannot be easily extended to reputation systems supporting multi-nominal rating levels.

Second, most of the existing approaches do not have all the capabilities listed in Table 2.3. For example, the iterated filtering approach can work when personal ratings are lacking, but it cannot work when majority ratings are unfair. The personalized approach can work when majority ratings are unfair, but it cannot work when the witnesses change their behaviors. The approaches using only local information, such as the iterated filtering approach and entropy-based filtering approach, have the capability of working when witness behaviors change. But these approaches usually cannot handle the situation where majority ratings are unfair or the consumer's personal ratings are missing. The approaches using global information, such as TRAVOS and the personalized approach, are capable of working when majority ratings are unfair, but they cannot work when witness behaviors vary over providers.

Therefore, the research work in this thesis focuses on designing approaches having all

the capabilities by considering personal, public, local, and global information together to build reliable reputation systems and secure online service provision environment. In our current work, we do not explicitly describe what reputation system architecture the approach is fit into. We make an assumption that if a centralized architecture is adopted, the providers' ratings are shared among all consumers. If a decentralized architecture is adopted, we assume that the consumer can obtain the testimonies from the witnesses without discussing how the consumer discovers the witnesses and collects testimonies. Though we have mentioned that continuous ratings provide the richest power in discriminating providers' trustworthiness, the use of continuous ratings requires extra overhead to store the rating information. We currently focus our work on addressing the problem of unfair testimonies for reputation systems accepting discrete rating levels. In addition, we also make an assumption that the consumers in the reputation system usually have multiple transactions with a provider to ensure the accuracy of identifying dishonest witnesses and evaluating provider reputation.

Chapter 3

A Two-Stage Clustering Filtering Approach

In this chapter, we first present a two-stage clustering filtering approach¹ to address the problem of unfair testimonies for reputation systems supporting multi-nominal rating levels. Then we simulate an online service provision environment to explore the performance of the proposed two-stage clustering filtering approach in various scenarios.

3.1 Introduction

As we have pointed out, mitigating the influence of unfair testimonies is one important requirement to build a reliable reputation system. Nevertheless, according to what we have reviewed in Chapter 2, most of the existing approaches only support reputation systems accepting binary ratings. To address this limitation, in this chapter we propose a two-stage clustering filtering approach to filter unfair testimonies for reputation systems supporting multi-nominal rating levels.

Before we present the proposed approach, we need to clarify what “*unfair testimonies*” means in our context. According to the definition of trust [21] we introduced in Chap-

---

¹This work is partially published in AAMAS2010 (conference paper [6] in author’s publication list) and ICCT2011 (conference paper [5] in author’s publication list).

ter 1: “a particular level of the subjective probability with which an agent assesses that another agent or another group of agents will perform a particular action, both before he can monitor such action and in a context in which it affects his own actions”, the trustworthiness of the target provider P is an opinion (i.e., a particular level of the subjective probability) held by the consume U towards P . Therefore, we consider that whether a witness’s testimonies are unfair (or whether the witness is dishonest in reporting testimonies) should also be the opinion held by U towards the witness. An intuition is that the more similar to U ’s past personal ratings regarding P the witness’s testimonies are, the more likely the testimonies are fair. Hence, if we can group with U the witnesses that provide similar testimonies as U ’s personal ratings, we can actually find the honest witnesses and filter the unfair testimonies. We need to emphasize that “*unfair testimonies*” does not always mean that the witnesses report ratings intentionally unfairly. The “*unfairness*” can be due to the subjective difference. In this chapter, we still consider the different testimonies caused by subjective difference as unfair.

To group the similar testimonies together, the technique of clustering is a good choice. The basic idea of the proposed two-stage clustering filtering approach is as follows. Imagine the accumulated ratings from a particular witness or the consumer himself as a feature vector in a K -dimensional space (K is the number of rating levels used in the reputation system). In this chapter, we introduce a two-stage clustering filtering approach to group the feature vectors with close similarity into one cluster. After clustering, the testimonies will be discarded if they are not in the cluster including the consumer’s personal rating vector (if any), or if they are not in the cluster including the majority witnesses’ rating vectors.

This chapter is organized as follows. Section 3.2 introduces the notations we use in this chapter. Section 3.3 presents the proposed two-stage clustering filtering approach. Experimental results and discussion regarding the results are presented in Section 3.4.

Finally, Section 3.5 summarizes this chapter.

3.2 Notations

Before getting into the details of the proposed two-stage clustering filtering approach, we first introduce the notations we use in this chapter.

3.2.1 Aggregating Consumers' Ratings

Suppose that in a reputation system, there are M providers $\{P_1, P_2, \dots, P_M\}$, and N consumers $\{U_1, U_2, \dots, U_N\}$. After a transaction between a consumer U_n ($1 \leq n \leq N$) and a provider P_m ($1 \leq m \leq M$) is completed at time t , U_n can rate P_m 's behavior as a rating level from a set of predefined discrete rating levels. Suppose that there are K ($K \geq 2$) different rating levels which are indexed by $1, 2, \dots, K$. If U_n rates P_m 's behavior as rating level k , U_n 's rating for P_m is represented as a row vector:

$$r_{U_n, P_m}^t = [r_{U_n, P_m}^t(1), \dots, r_{U_n, P_m}^t(i), \dots, r_{U_n, P_m}^t(K)], \quad (3.1)$$

where $r_{U_n, P_m}^t(k) = 1$ and $r_{U_n, P_m}^t(i) = 0$ ($1 \leq i \leq K$ and $i \neq k$). For example, suppose that $K = 5$ and U_n rates P_m 's behavior as 4 after one transaction, then $r_{U_n, P_m}^t = [0, 0, 0, 1, 0]$. Suppose that we take the ratings provided in the recent F time windows into account (e.g., in eBay, a seller's reputation can be calculated based on the ratings within recent 3 months). In a single time window f , the aggregated ratings from U_n for P_m can be represented as a row vector R_{U_n, P_m}^f , expressed as:

$$R_{U_n, P_m}^f = [R_{U_n, P_m}^f(1), \dots, R_{U_n, P_m}^f(i), \dots, R_{U_n, P_m}^f(K)], \quad (3.2)$$

where $R_{U_n, P_m}^f(i)$ is the aggregated result of $r_{U_n, P_m}^t(i)$ ($i = 1, \dots, K$ and t is in time window f). The updating of R_{U_n, P_m}^f can be achieved by adding the new rating vector r_{U_n, P_m}^t to the previous aggregated rating vector R_{U_n, P_m}^f ².

3.2.2 Aggregating Ratings with Aging

Providers may change their behaviors over time. Therefore, P_m 's more recent behavior should be assigned a greater weight, and his older behavior should be forgotten continuously. To achieve this, a time forgetting factor λ ($\lambda \in [0, 1]$) is introduced in [1] [59] to control the rate at which P_m 's old behavior is forgotten. In the proposed two-stage clustering filtering approach, we also adopt this strategy. When $\lambda = 0$, P_m 's old behavior is completely forgotten after a single time window elapses. When $\lambda = 1$, P_m 's old behavior is never forgotten. Let P_m 's accumulated ratings with aging from U_n after time window f be denoted as:

$$A_{U_n, P_m}^f = [A_{U_n, P_m}^f(1), \dots, A_{U_n, P_m}^f(i), \dots, A_{U_n, P_m}^f(K)]. \quad (3.3)$$

Then the accumulated rating vector with aging after time window f can be calculated as:

$$\begin{cases} A_{U_n, P_m}^0 = \vec{0}, \\ A_{U_n, P_m}^f = \lambda \times A_{U_n, P_m}^{f-1} + R_{U_n, P_m}^f (f \in \mathbb{N}). \end{cases} \quad (3.4)$$

When U_n is evaluating P_m 's reputation, he can collect rating vectors from other consumers to facilitate his evaluation. Then the set of the consumers who provide rating vectors to U_n regarding P_m is expressed as:

$$W_{U_n, P_m} = \{U_j \mid j \neq n \wedge \|A_{U_j, P_m}^f\| \neq 0\}. \quad (3.5)$$

---

²The adding mentioned here is referred to as a matrix addition.

From U_n 's point of view, W_{U_n, P_m} is called the set of witnesses regarding P_m (each consumer in W_{U_n, P_m} is a witness), and the rating vector provided by each witness is called testimony from this witness.

The local information regarding P_m until time window f is denoted as:

$$L_{U_n, P_m}^f = \{A_{U_j, P_m}^f \mid U_j \in W_{U_n, P_m} \cup \{U_n\} \wedge \|A_{U_j, P_m}^f\| \neq 0\}. \quad (3.6)$$

Then U_n can use L_{U_n, P_m}^f to evaluate P_m 's reputation by applying some existing reputation systems, such as BRS [39] and Dirichlet Reputation System (DRS) [59]. But as we have pointed out, though U_n can use the testimonies to facilitate his evaluation regarding P_m 's reputation, the testimonies may mislead U_n 's evaluation if the witnesses do not provide testimonies in an honest way. This may even result in an opposite evaluation situation, for example, where a very low reputation is evaluated regarding a reputable provider. Therefore, we need to filter the testimonies provided by dishonest witnesses before aggregating them together with the testimonies provided by honest witnesses.

3.3 The Proposed Two-Stage Clustering Filtering Approach

Clustering is originally used to assign a set of observations into subsets (called clusters) so that the observations in the same cluster are similar to each other according to some criteria [67]. There are many clustering methods designed, such as k -means clustering method and hierarchical clustering method. For the proposed two-stage clustering filtering approach, we use hierarchical clustering method as it supports different clustering stopping criteria. For hierarchical clustering method, one observation is initially considered as one cluster, then two clusters will be merged together according to some distance criteria

(e.g., merge two clusters together which have shortest distance between each other). The merging process continues until the predefined clustering stopping criteria (e.g. predefined fixed number of clusters) is met.

Here we propose a two-stage clustering filtering approach to filter unfair testimonies. In the first stage, rating vectors are grouped into a predefined number (κ) of clusters. More specifically, initially each rating vector reported by each witness is considered as a cluster, and then each two clusters with the shortest 2-norm distance will be merged into a new cluster. After that, each two clusters with the shortest 2-norm distance will again be merged into a new cluster. The process continues until the predefined cluster number κ is met (we will show how κ value will impact the accuracy of the two-stage clustering filtering approach in Section 3.4). The stage aims to merge together rating vectors that have the closest similarity.

However, if the predefined κ is too large, some similar vectors may not successfully be clustered together as expected in Stage 1. Thus, a second stages is employed. In this stage, the merging process will continue until a different criteria is satisfied. More specifically, in this stage, the approach first computes the furthest distance between any two clusters produced in Stage 1, then merges two clusters with the minimum furthest distance if this distance is smaller than the predefined distance threshold (σ). The merging process will stop when no furthest distance between any two clusters is smaller than σ . Stage 2 aims to merge together rating vectors based on their similarities, while ensuring that vectors that are obviously different are not merged by properly setting σ value (we will show how σ value will impact the accuracy of the two-stage clustering filtering approach in Section 3.4). The detailed algorithm is presented in Algorithm 1.

In Algorithm 1, the two-stage clustering filtering approach first finds the witnesses and collects the testimonies regarding the target provider P (see Line 1-5). As we have mentioned in Chapter 2, we do not discuss here how to discover the witnesses and testimonies

---

Procedure: Two-Stage Clustering Filtering

Input : P , the provider whose reputation is being evaluated;
 U , the consumer who is evaluating P 's reputation;
 κ , the threshold controlling stage 1 clustering stopping;
 σ : the threshold controlling stage 2 clustering stopping

Output : a set of honest witnesses regarding P ;

```

1 Find  $W_{U,P}$  and collect rating vectors from  $W_{U,P}$ ;
2 if  $\|A_{U,P}\| \neq 0$  then
3   |  $W'_{U,P} = W_{U,P} \cup \{U\}$ ;
4 else
5   |  $W'_{U,P} = W_{U,P}$ ;
6 Normalize each rating vector  $A_{U_j,P}$  ( $U_j \in W'_{U,P}$ ) as  $A'_{U_j,P}$ ;
7 foreach  $U_j \in W'_{U,P}$  do
8   |  $C_j = \{R'_{U_j,P}\}$ ;
9  $C = \bigcup \{C_j\}$ ;
10 while  $|C| > \kappa$  do
11   | foreach  $C_i \in C, C_j \in C$  ( $i \neq j$ ) do
12     | foreach  $A'_{U_{i'},P} \in C_i, A'_{U_{j'},P} \in C_j$  do
13       |  $d_s(i', j') = \|A'_{U_{i'},P} - A'_{U_{j'},P}\|$ ;
14       |  $D_s(C_i, C_j) = \min(d_s(i', j'))$ ;
15     |  $p, q = \arg\{\min_{i,j} D_s(C_i, C_j)\}$ ;
16     | merge  $C_p$  and  $C_q$  together;
17  $stopcondition = \text{false}$ ;
18 while  $stopcondition$  is false do
19   | foreach  $C_i \in C, C_j \in C$  ( $i \neq j$ ) do
20     | foreach  $A'_{U_{i'},P} \in C_i, A'_{U_{j'},P} \in C_j$  do
21       |  $d_s(i', j') = \|A'_{U_{i'},P} - A'_{U_{j'},P}\|$ ;
22       |  $D_f(C_i, C_j) = \max(d_s(i', j'))$ ;
23     |  $p, q = \arg\{\min_{i,j} D_f(C_i, C_j)\}$ ;
24     | if  $D_f(C_p, C_q) \leq \sigma$  then
25       | merge  $C_p$  and  $C_q$  together;
26     | else
27       |  $stopcondition = \text{true}$ ;
28 if  $\|A_{U,P}\| > 0$  then
29   |  $\exists u, A'_{U,P} \in C_u$ ;
30   | return  $W_T = \{U_j | A'_{U_j,P} \in C_u\}$  as honest witnesses;
31 else
32   |  $q = \arg\{\max_i |C_i|\}$ 
33   | return  $W_T = \{U_j | A'_{U_j,P} \in C_q\}$  as honest witnesses;

```

Algorithm 1: Two-stage clustering filtering approach

in a decentralized reputation system. We assume that the consumer can always obtain testimonies in some way. For simplicity, we use $A_{U,P}$ to denote the accumulated aged rating vector from U regarding P till the moment when U is evaluating P 's reputation. Before clustering, the rating vectors need to be normalized (Line 6). The normalization for each rating vector is achieved by having the value of each dimension divided by the sum of the value of each dimension (e.g., suppose that a rating vector is $[0, 1, 4, 2, 1]$, then the normalized rating vector is $[0, 0.125, 0.5, 0.25, 0.125]$). After normalization, the clustering process begins. Lines 7-16 are for Stage 1 clustering. Initially, an individual normalized rating vector is considered as a cluster (Lines 7-9). Then the shortest distance between any two clusters³ is computed as Lines 11-14 show. The two clusters whose shortest distance is minimum are merged together (Lines 15-16). The process continues until the number of the clusters is smaller than or equal to κ . Lines 17-27 are for Stage 2 clustering. In this stage, the furthest distance⁴ is computed for any two clusters produced in the first stage as Lines 19-22 show. Then the two clusters whose furthest distance is minimum are merged together (Lines 23-25). The process continues until the furthest distance between any two clusters is larger than σ (Lines 26-27). After we get the clustering results, there are two scenarios for consideration as described below:

1. If the consumer U has personal ratings for the target provider P , then the witnesses whose rating vectors are included in the cluster containing the rating vector from U are returned as honest witnesses (Lines 28-30).
2. If the consumer U does not have personal ratings for the target provider P , then the witnesses whose rating vectors are included in the cluster containing the greatest number of rating vectors are returned as honest witnesses (Lines 31-33). It is worth

---

³If either or both clusters include more than one normalized rating vectors, the shortest distance is calculated as the minimum of the distance of any two rating vectors, one of which is from one cluster and the other one is from the other cluster.

⁴The similar method as in Stage 1 clustering is adopted if either or both clusters include more than one normalized rating vector.

pointing out that there are other options to select a particular cluster. In our current experiments, we follow the majority rule to select the cluster including the greatest number of rating vectors. Therefore, it will lead to inappropriate cluster selected in the scenario where majority testimonies are unfair. Other options can be, for example, making a decision by incorporating other information, such as global information.

It needs to point out that the two-stage clustering filtering approach is proposed for filtering unfair testimonies. To evaluate the provider's reputation, the proposed approach needs to be integrated with a reputation evaluation mechanism, such as BRS [39] and DRS [59], to get the provider's reputation by passing the achieved honest witnesses' testimonies to the reputation system.

3.4 Experimental Studies

In current research on coping with the problem of unfair testimonies, it has always been a difficulty to obtain real-world data to conduct experiments due to the following three reasons. First, it is difficult to obtain the real-world data for an online service provision system due to the privacy considerations by the business owners. Second, even if we can get the data from an online service provision system (e.g., eBay), there is usually no ground truth based on which testimonies can be labeled as unfair. Third, it is difficult to analyze the influence of different strategies of inserting unfair testimonies because of the opaque of ground truth. Therefore, most of current research works are using simulated synthetic data to conduct experiments. According to the analysis of 71 research papers on the topic of trust management published in the 3 top conferences in agent area – Association for the Advancement of Artificial Intelligence Conference (AAAI), the International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), and the International Joint

Conference on Artificial Intelligence (IJCAI) – from 2002 to 2012, it is found that over 80% of the papers use simulations to assess the performance of their proposed trust models. Hence, in this thesis work, we will also use the simulated synthetic data to conduct experiments.

We carry out three sets of experiments to evaluate the accuracy and robustness of the two-stage clustering filtering approach in various scenarios. The first set investigates the relationship between the threshold values and the accuracy of the two-stage clustering filtering approach in filtering unfair testimonies. The second set examines the robustness of the two-stage clustering filtering approach in various scenarios where multi-nominal rating levels are adopted. The third set compares the two-stage clustering filtering approach with other two representative approaches – iterated filtering approach [1] and TRAVOS [48] – in terms of the accuracy of filtering unfair testimonies and provider reputation evaluation when only binary ratings are accepted.

We measure the accuracy of an approach in filtering unfair testimonies in two ways. First, the accuracy can be measured by the capability of the approach in detecting dishonest witnesses, which can be more specifically measured by the false positive rate (FPR) and false negative rate (FNR), computed as:

$$FPR = \frac{f_p}{f_p + t_n} \quad (3.7)$$

$$FNR = \frac{f_n}{f_n + t_p} \quad (3.8)$$

where f_p , t_p , f_n and t_n represents the number of false positives, true positives, false negatives and true negatives, respectively. In our experiments, a true positive means that an honest witness is correctly detected as honest; a false positive means that a dishonest one is incorrectly detected as honest; a true negative means that a dishonest one is correctly filtered as dishonest; a false negative means that an honest witness is

incorrectly filtered as dishonest. Lower FPR and FNR values imply higher accuracy.

The second way is to use Matthew's Correlation Coefficient (MCC) [68] to generally measure an approach's filtering accuracy, which is computed as:

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

MCC value is between -1 and 1, where 1 represents a perfect filtering result, -1 represents an inverse filtering result, and 0 represents a random filtering result.

Our experiments involve two types of dishonest witnesses [45]:

- *ballot-stuffing* witnesses report testimonies that the provider behaves well regardless of the true behavior of the provider;
- *badmouthing* witnesses report testimonies that the provider behaves poorly regardless of the true behavior of the provider.

3.4.1 Threshold Value Exploration

In this set of experiments, we explore how the threshold values (κ and σ) impact the accuracy of the two-stage clustering filtering approach in filtering unfair testimonies and try to establish an optimum threshold value range. We simulate an online service provision environment which includes 1 provider P , ω witnesses, and 1 consumer U . The provider P has an initial willingness (iw) value selected from the value set $\{0.1, 0.3, 0.5, 0.7, 0.9\}$. Each witness (or U) has T transactions with P . For each transaction, P 's rating is simulated by a willingness value which is generated from a normal distribution whose mean is equal to iw , and standard deviation is δ . The mapping between the generated willingness value and the rating level (5 rating levels are simulated) for P is shown in Table 3.1.

Willingness	$(-\infty, 0.2]$	$(0.2, 0.4]$	$(0.4, 0.6]$	$(0.6, 0.8]$	$(0.8, \infty)$
Rating	1	2	3	4	5

Table 3.1: Mapping from willingness value to rating level

Each dishonest witness will report ratings unfairly with a randomly generated probability value which is larger than r . The percentage of ballot-stuffing witnesses and badmouthing witnesses is p_h and p_l , respectively. We assume that when P 's iw value is very low (i.e., $iw = 0.1$ or $iw = 0.3$), there are no badmouthing witnesses. When P 's iw value is very high (i.e., $iw = 0.7$ or $iw = 0.9$), there are no ballot-stuffing witnesses.

We explore how the threshold values will impact the accuracy of the two-stage clustering filtering approach from two aspects – scalability and stability. Here the scalability means the number of witnesses reporting ratings to the consumer, and the stability means the number of transactions happening between each witness (or U) and P . A high scalability means that there are a lot of witnesses. On the contrary, a low scalability means that there are few witnesses. A high stability means that there are a large number of transactions between P and the witnesses (or U), and a low stability means that there are only a few transactions between P and the witnesses (or U). We explore two levels of scalability and stability – high (i.e., 100 witnesses and 100 transactions) and low (i.e., 10 witnesses and 10 transactions). Table 3.2 lists the parameter meanings and possible values we adopt in our simulation.

In our simulation, we set the forgetting factor as $\lambda = 0.9$. κ increases from 1 to ω with step 1. σ increases from 0.1 to 1.4 (the largest 2-norm distance value between two normalized rating vectors) with step 0.1. We have 40 parameter value combinations. For each parameter value combination, there are two scenarios: U has transactions with P , and U has no transactions with P . Therefore, we have in total 80 simulation scenarios. We run 100 rounds for each simulation scenario to achieve a statistical accuracy. As an illustration, we show the results of the MCC, FPR, and FNR value changes with κ and ω

Parameter	Meaning	Value
ω	The number of witnesses	$\{10,100\}$
iw	P 's initial willingness value	$\{0.1,0.3,0.5,0.7,0.9\}$
δ	The standard deviation of the normal distribution to simulate P 's behavior	$\{0.1,0.2\}$
T	The number of transactions between each witness or U and P	$\{10,100\}$
p_h	The percentage of ballot-stuffing witnesses	$\{0, 15\%, 30\%\}$
p_l	The percentage of badmouthing witnesses	$\{0, 15\%, 30\%\}$
r	The smallest value of the probability of the dishonest witnesses reporting ratings unfairly	0.5

Table 3.2: Simulation parameters, meanings and values

values in the scenario where $iw = 0.5$, $\delta = 0.2$, $p_h = 15\%$ and $p_l = 15\%$ when scalability and stability are high or low, and U has no transactions with P . Then we have four sets of results as shown in Figure 3.1 and 3.2.

Figure 3.1 shows the MCC, FPR and FNR values when the scalability is high (i.e., $\omega = 100$). According to the results, when the scalability is high, MCC (FPR or FNR) value presents a similar changing trend when the stability is high or low. MCC value first increases with σ value when $\kappa > 10$, then it keeps stable for some σ values, and then decreases until the largest σ value is reached. When $\kappa \leq 10$, it can be noticed that when $\sigma < 1.1$, the two-stage clustering filtering approach can achieve a good MCC result.

To investigate the results more closely, Figure 3.1(c) and (d) show FPR values when stability is high and low, respectively. Figure 3.1(e) and (f) show FNR values when stability is high and low, respectively. We can see that when σ value is very large (i.e., 1.2-1.4), FPR is very high. As when σ value is very large, the clusters achieved from the first stage will be wrongly merged together in the second stage. More specifically, the clusters including the dishonest witnesses are merged with the clusters including the

(a)

(b)

(c)

(d)

(e)

(f)

Figure 3.1: (a) MCC value when $T=100$ and $\omega = 100$; (b) MCC value when $T=10$ and $\omega = 100$; (c) FPR value when $T=100$ and $\omega = 100$; (d) FPR value when $T=10$ and $\omega = 100$; (e) FNR value when $T=100$ and $\omega = 100$; (f) FNR value when $T=10$ and $\omega = 100$

(a)

(b)

(c)

(d)

(e)

(f)

Figure 3.2: (a) MCC value when $T=100$ and $\omega = 10$; (b) MCC value when $T=10$ and $\omega = 10$; (c) FPR value when $T=100$ and $\omega = 10$; (d) FPR value when $T=10$ and $\omega = 10$; (e) FNR value when $T=100$ and $\omega = 10$; (f) FNR value when $T=10$ and $\omega = 10$

honest witnesses, leading to a high FPR value. For FNR value, it can be noticed that FNR is very high when κ value is large (i.e, 40-100) and σ value is small (i.e., 0.1-0.7). The reason is that when κ value is large, the two-stage clustering filtering approach will finish the first stage clustering quite early, leaving some clusters including honest witnesses alone. But at the second stage, the clusters including the honest witnesses cannot be merged together as σ value is too small, finally leading to a high FNR value. Generally speaking, when σ value is too large, FPR will be very high, and when κ value is too large and σ value is too small, FNR will be very high. A noticed trend is that FNR value increases with the value of $\frac{\kappa}{\omega}$ increasing when σ value is small.

Figure 3.2 shows MCC, FPR and FNR values when the scalability is low (i.e., $\omega = 10$). It can be noticed that MCC (FPR or FNR) values show a similar changing trend as those shown in Figure 3.1 when the stability is high or low.

As a summary, a larger σ value will cause a higher FPR no matter the scalability is high or low. And a smaller σ value will cause a higher FNR when $\frac{\kappa}{\omega}$ is large. Therefore, before choosing the appropriate threshold values, the consumer better evaluates the scalability and stability of the collected testimonies. If the consumer cares more about FPR, then a smaller σ value is a good choice though this may lead to a higher FNR. If the consumer cares more about FNR, then a smaller κ value is a better choice. As a balance, it is a good practice to adopt a relatively small κ value (≤ 10) and a medium σ value (0.7-0.9). In our following experiments, we generally explore the accuracy of the two-stage clustering filtering approach in the scenarios when the scalability is medium (i.e., $\omega = 20$) and stability is randomly generated. The threshold values we use in the following simulation are set as $\kappa = 10$ and $\sigma = 0.7$.

3.4.2 Robustness Investigation

The aim of this set of experiments is to investigate the robustness of the two-stage clustering filtering approach in various scenarios when appropriate threshold values are set. In this set of experiments, we simulate a similar service provision environment as in the first set of experiments, including 1 provider P , 20 witnesses and 1 consumer U . The standard deviation of the normal distribution to simulate the provider's behavior is 0.2. P 's rating of each transaction is generated through the same way as in the first set of experiments.

3.4.2.1 Dishonest Witness Percentage and Transaction Number Varying

In this experiment, we first explore the accuracy of the two-stage clustering filtering approach when the percentage of the dishonest witnesses changes. Figure 3.3 shows the MCC value changes with the percentage of the dishonest witnesses. When P 's iw value is 0.1 or 0.3, $p_l = 0$, and p_h increases from 10% to 90%. When P 's iw value is 0.7 or 0.9, $p_h = 0$, and p_l increases from 10% to 90%. When P 's iw value is 0.5, $p_h = 20\%$, and p_l increases from 0 to 70%. Each witness (or U) will have a randomly generated number of transactions (in the range of $[10,100]$) with P .

Figure 3.3: (a) U has no transactions; (b) U has transactions

According to the results, when U has no transactions, MCC value is close to 1 until the dishonest witnesses occupy as the majority. For example, when P 's iw value is 0.5 and p_l is smaller than 40%, MCC value is close to 1. As the percentage of the honest witnesses is greater than 40% which is greater than p_l or p_h , the two-stage clustering filtering approach will follow the majority rule and a correct filtering result can be achieved. But when p_l is greater than 40%, the percentage of the honest witnesses is smaller than p_l . In this case, the two-stage clustering filtering approach will return the badmouthing witnesses as honest witnesses by following the majority rule. When U has transactions with P , however, the two-stage clustering filtering approach can work properly even when the dishonest witnesses occupy as the majority according to the results shown in Figure 3.3(b) (note that the y-axis value starts from 0.85 in Figure 3.3(b)). MCC value is close to 1 even when the percentage of the dishonest witnesses is 90%. As when U has transactions with P , the two-stage clustering filtering approach will return the witnesses that have the similar rating vectors as U 's personal rating vector as honest witnesses. Assuming that the honest witnesses should have similar testimonies as U 's personal ratings, the two-stage clustering filtering approach can get a correct filtering result even when the dishonest witnesses occupy as the majority.

Second, in this experiment, we explore the accuracy of the two-stage clustering filtering approach changes with the number of the transactions between each witness and P (for simplicity, U has no transactions with P because when the dishonest witnesses do not occupy as the majority, the filtering result when U has no transactions is similar to that when U has transactions). When P 's iw value is 0.1 or 0.3, $p_l = 0$ and $p_h = 30\%$. When P 's iw value is 0.7 or 0.9, $p_l = 30\%$ and $p_h = 0$. When P 's iw value is 0.5, $p_l = 15\%$ and $p_h = 15\%$. The number of the transactions between each witness and P increases from 1 to 50. The results are shown in Figure 3.4. We can see that MCC value increases with the number of transactions. Initially, MCC value is very low. After about 10 transactions,

MCC value approaches to 1, meaning that the two-stage clustering filtering approach performs more and more stably.

Figure 3.4: MCC value changes with the number of transactions

3.4.2.2 Provider Behavior Varying

In this experiment, we investigate the robustness of the two-stage clustering filtering approach in the scenarios where P 's behavior changes over time. As P 's behavior is simulated by the normal distribution, we change the mean of the normal distribution to simulate his behavior change. In this experiment, we simulate four scenarios: P 's iw value increases from 0.1 to 0.3 then to 0.5; P 's iw value decreases from 0.5 to 0.3 then to 0.1; P 's iw value decreases from 0.9 to 0.7 then to 0.5; P 's iw value increases from 0.5 to 0.7 then to 0.9. In the first two scenarios, $p_h = 30\%$ and $p_l = 0$. In the last two scenarios, $p_l = 30\%$ and $p_h = 0$. For simplicity, U has no transactions with P . Figure 3.5 shows the MCC value changes with the number of transactions. According to the results, MCC value is initially very low as the number of the transactions is small. Then after about 15 transactions, the two-stage clustering filtering approach performs more stably and MCC value approaches to 1. Though after each 20 or 40 transactions, the provider changes his

behavior, the two-stage clustering filtering approach still can work properly and stably.

Figure 3.5: (a) P changes his behavior after each 20 transactions; (b) P changes his behavior after each 40 transactions

3.4.3 Comparative Experiments

We compare the proposed two-stage clustering filtering approach with other approaches from two aspects – the accuracy of filtering accuracy and evaluating provider reputation. Though the motivation for the two-stage clustering filtering approach is to address the problem of unfair testimonies for reputation systems supporting multi-nominal rating levels, it can also work for reputation systems accepting binary rating levels. In this experiment, we compare the two-stage clustering filtering approach with the iterated filtering approach [1] in terms of filtering accuracy. We also compare the two-stage clustering filtering approach with both the iterated filtering approach and TRAVOS [48] in terms of the accuracy of provider reputation evaluation. Note that in this set of experiments, after filtering unfair testimonies by the two-stage clustering approach or the iterated filtering approach, the achieved fair testimonies will be aggregated to evaluate provider's reputation using BRS [39].

In this set of experiments, we simulate a similar service provision environment as that

simulated in the second set of experiments, which includes 1 provider P , 20 witnesses, and 1 consumer U . For each transaction, a willingness value is generated for P , which indicates in what probability P 's behavior will be rated as positive. The willingness value for the first transaction is equal to P 's initial willingness value (iw) from the value set $\{0.1, 0.2, 0.3, 0.7, 0.8, 0.9\}$ ⁵. For the subsequent transactions, the willingness value is randomly generated through one of the following three strategies – the willingness value of last transaction adding 0.02, equal to the willingness value of last transaction, or the willingness value of last transaction subtracting 0.02. The willingness value for each transaction is also limited in the range of $[iw - 0.1, iw + 0.1]$. In this experiment, one type of dishonest witnesses is studied, who will report the opposite of the actual ratings as testimonies. When comparing with TRAVOS, U will have transactions with P as TRAVOS needs the consumer's personal transactions to work.

Figure 3.6 shows the MCC value changes with the percentage of dishonest witnesses. The number of the transactions between each witness (or U) and P is 20. As Figure 3.6(a) shows, the two-stage clustering filtering approach can filter the dishonest witnesses when the dishonest witnesses do not occupy as the majority when U has no transactions. When U has transactions, the two-stage clustering filtering approach can filter the unfair testimonies no matter what the percentage of the dishonest witnesses is. The MCC value using the iterated filtering approach continuously decreases with the percentage of dishonest witnesses. It can be noticed that when there are 30% dishonest witnesses, MCC using the iterated filtering approach is close to 0.5, implying that using the iterated filtering approach is close to random guessing.

Figure 3.7 shows the evaluated provider reputation changes with the percentage of dishonest witnesses increasing after using different approaches. According to the results,

---

⁵If P 's initial willingness is 0.4, 0.5, or 0.6, the ratings from the dishonest witnesses we study in the experiments are quite similar to the ratings from the honest witnesses. Therefore, we do not simulate the scenarios where P 's iw value is 0.4, 0.5 or 0.6.

Figure 3.6: (a) U has no transactions; (b) U has transactions

the reputation value by using the two-stage clustering filtering approach is very close to the expected reputation value (expected reputation value is calculated using the actually fair testimonies only). The reputation value after using the iterated filtering approach is initially very close to the expected reputation value, then continuously deviates from the expected reputation value with the percentage of dishonest witnesses increasing. TRAVOS works stably but its result is not as close to the expected reputation value as that by using the two-stage clustering filtering approach.

Figure 3.7: (a) $iw = 0.1$; (b) $iw = 0.9$

Figure 3.8 shows the evaluated provider reputation changes with the number transac-

tions after using different approaches. In Figure 3.8(a) and (b), the percentage of dishonest witnesses is 20%, and iw value is 0.1 and 0.9, respectively. In Figure 3.8(c) and (d), the percentage of dishonest witnesses is 90%, and iw value is 0.1 and 0.9, respectively.

Figure 3.8: (a) $iw = 0.1$ and 20% dishonest witnesses; (b) $iw = 0.9$ and 20% dishonest witnesses; (c) $iw = 0.1$ and 90% dishonest witnesses; (d) $iw = 0.9$ and 90% dishonest witnesses

According to the results, when the percentage of the dishonest witnesses is 20%, the reputation value after using the two-stage clustering filtering approach is very close to the expected reputation value after about 10 transactions. The reputation value after using the iterated filtering approach is close to the expected reputation value after about 10 transactions, but then deviates from the expected reputation value after about 30 transactions. The TRAVOS reputation value continuously approaches to the expected reputation

value with the number of transactions increasing but it needs more transactions. When the percentage of the dishonest witnesses is 90%, the reputation value after using the two-stage clustering filtering approach is still very close to the expected reputation value. But the reputation value after using the iterated filtering approach seriously deviates from the expected reputation value, and is even worse than the result of not using any filtering approach. The reputation value after using TRAVOS continuously approaches to the expected reputation value but it needs a longer time.

3.4.4 Discussions on Experimental Results

We have conducted three sets of experiments in this section. In the first set of experiments, we explored the influence of threshold values on the accuracy of the two-stage clustering filtering approach when different levels of scalability and stability are set. According to the results, a larger σ value will lead to a higher FPR, a smaller σ value and a larger $\frac{\kappa}{\omega}$ value will lead to a higher FNR. Therefore, a medium σ value and a smaller $\frac{\kappa}{\omega}$ value may be more appropriate for a general application. For different systems, it is better to consider current scalability and stability levels and set the threshold values according to the FPR and FNR concerns.

In the second set of experiments, we investigated the robustness of the two-stage clustering filtering approach in various scenarios when multi-nominal rating levels are adopted. When the witnesses or the consumer has more than 10-15 transactions with the provider, the two-stage clustering filtering approach can work more stably and accurately in filtering unfair testimonies. Therefore, it also suggests that when the consumer is collecting testimonies from the witnesses, it is better to consider the testimonies from the witnesses that have more than 10-15 transactions with the provider. When the consumer has no transactions with the provider, the two-stage clustering filtering approach can

work accurately until the dishonest witnesses occupy as the majority. When the consumer has transactions with the provider, the two-stage clustering filtering approach can work properly no matter what the percentage of the dishonest witnesses is. In the scenarios where the provider's behavior changes over time, the provider behavior change has no obvious influence on the accuracy of the two-stage clustering filtering approach working. It still works more stably and accurately after about 15 transactions.

In the third set of experiments, we compared the two-stage clustering filtering approach with the iterated filtering approach and TRAVOS from two aspects – the accuracy of filtering and evaluating provider reputation. Compared with the iterated filtering approach, the two-stage clustering filtering approach performs better on the accuracy of filtering unfair testimonies, and hence achieves a more accurate result on provider reputation evaluation. Compared with TRAVOS, the two-stage clustering filtering approach can accurately evaluate the provider's reputation at a faster rate.

3.5 Summary

In this chapter, we proposed a two-stage clustering filtering approach to filter unfair testimonies for reputation systems using multi-nominal ratings. The two-stage clustering filtering approach incorporates local information, public information and personal information (when available), and adopts clustering technique to filter unfair testimonies for reputation systems supporting multi-nominal rating levels. As the experimental results show, the approach can effectively filter unfair testimonies in various scenarios when appropriate threshold values are set.

The two-stage clustering filtering approach has the following capabilities. First, as it considers personal information besides local information and public information, it can work when dishonest witnesses occupy as majority as it assigns the personal information

with a higher priority (i.e., the witnesses who have the similar rating vectors as the consumer's personal rating vector will be considered as honest even though they may not have the similar rating vectors as majority witnesses) when personal information is available. Second, it can resist flooding attack because the collected testimonies (i.e., rating vectors) is normalized first before clustering. Third, it can work when the consumer has no personal ratings. But as the two-stage clustering filtering approach currently follows the majority rule, it cannot address the scenario where the consumer has no personal ratings and at the same time a particular type of witnesses occupy as the majority. Fourth, it can cope with the problem of provider behavior varying by employing forgetting factor. Fifth, it will not be influenced by witness behavior change over providers as only the testimonies regarding the target provider is considered. In addition, as simulated in the first and second sets of experiments where badmouthing and ballot-stuffing witnesses co-exist together, it also has the capability of coping with the situation that various types of dishonest testimonies exist at the same time.

Chapter 4

An Integrated Clustering Filtering Approach

In Chapter 3, we presented a two-stage clustering filtering approach. Though as the experimental results demonstrated in Chapter 3, the two-stage clustering filtering approach can filter the unfair testimonies with high accuracy and improve the reliability of reputation systems, especially the ones supporting multi-nominal ratings, there is still room for improvement. For example, the two-stage clustering filtering approach cannot work when a particular type of dishonest witnesses occupies as majority and coincidentally the consumer has no personal ratings regarding the target provider. To revamp the two-stage clustering filtering approach, we propose an integrated clustering filtering approach¹ in this chapter.

4.1 Introduction

In Chapter 3, a two-stage clustering filtering approach was proposed to tackle the problem of unfair testimonies in reputation systems. A limitation of the two-stage clustering filtering approach is that it has to follow the majority rule to filter unfair testimonies

---

¹This work is partially published in AAMAS2011 (conference paper [4] in author's publication list) and accepted for publication in *Computation Intelligence* (journal paper [1] in author's publication list).

when the consumer's personal ratings regarding the target provider are unavailable. As a consequence, it may not work when majority witnesses are dishonest.

To solve this problem, we propose an integrated *CLU*stering filtering approach (iCLUB) to filter unfair testimonies in this chapter. Compared to the two-stage clustering filtering approach using local, personal and public information to filter unfair testimonies, iCLUB also incorporates global information. More specifically, iCLUB adopts clustering technique and integrates two components, Local and Global. The Local component makes use of the consumer's knowledge about the target providers only. The Global component makes use of the consumer's knowledge about other providers that the witnesses have previously encountered. This is particularly useful when the consumer does not have much experience with the target provider.

The remainder of this chapter is structured as follows. Section 4.2 provides a brief description of the notations we use in this chapter. Section 4.3 presents the proposed iCLUB approach. Section 4.4 uses an example to illustrate how iCLUB works. The experimental studies and results are presented in Section 4.5. Finally, Section 4.6 summarizes this chapter.

4.2 Notations

Other than the notations defined in the previous chapter (see Section 3.2 in Chapter 3), we introduce some new notations here for the ease of presentation.

Compared to the two-stage clustering filtering approach, iCLUB leaves the problem of provider behavior varying to the reputation system it will be fit into. In each single time window f , iCLUB will be applied to filter the unfair testimonies in f . Therefore, the definition of local information is different from that defined in Chapter 3. Here, the

local information is defined as:

$$L_{U_n, P_m}^f = \{R_{U_j, P_m}^f \mid U_j \in W_{U_n, P_m} \cup \{U_n\} \wedge \|R_{U_j, P_m}^f\| \neq 0\}, \quad (4.1)$$

where U_n is the consumer who is evaluating the target provider P_m 's reputation, W_{U_n, P_m} is the set of witnesses providing testimonies to U_n regarding P_m as follows:

$$W_{U_n, P_m} = \{U_j \mid j \neq n \wedge \|R_{U_j, P_m}^f\| \neq 0\}, \quad (4.2)$$

where R_{U_j, P_m}^f is U_j 's aggregated rating vector in time window f (Equation (3.2)).

iCLUB uses local information and global information together to filter unfair testimonies. With the definition of local information, the global information G_{U_n, P_m}^f , which is the collection of each witness's local information, is expressed as:

$$G_{U_n, P_m}^f = \bigcup_{i=1, i \neq m}^M L_{U_n, P_i}^f. \quad (4.3)$$

4.3 The iCLUB Approach

In this section, we present the proposed iCLUB approach to filter unfair testimonies. In this work, we still do not differentiate the “unfairness” caused by malicious reporting and subjective difference. The iCLUB approach integrates two components – Local and Global. The Local component applies clustering on the local information regarding the target provider P , in the scenario where the consumer has a sufficient number of transactions with P . Otherwise, the Global component applies clustering on the global information. More details are given in the subsequent sections.

4.3.1 Making Use of Local Information

Considering the rating vector from the consumer or a particular witness as a feature vector in a K -dimensional space, the iCLUB approach groups the feature vectors with the same similarity into one cluster. After clustering, the rating vectors will be considered as unfair testimonies if they are not in the cluster which includes the consumer's personal rating vector. A pseudo code summary of this process is given in Algorithm 2.

Procedure: Local(P, U)

Input : P , the provider whose reputation is evaluated;
 U , the consumer evaluating P 's reputation;

Output : A set of honest witnesses regarding P ;

- 1 Collect local information regarding P as $L_{U,P}$;
- 2 $C_1, C_2, \dots, C_Z = \text{DBSCAN}(L_{U,P})$;
- 3 $\exists u, R_{U,P} \in C_u$ ($1 \leq u \leq Z$);
- 4 Return $W_T = \{U_i | R_{U_i,P} \in C_u \wedge U_i \neq U\}$;

Algorithm 2: Local component

The Local component of iCLUB first collects the local information regarding P (see Line 1)². DBSCAN [69] is then applied on the collected local information $L_{U,P}$ (for simplicity, we use $L_{U,P}$ to designate the local information in the time window iCLUB is applied to) to generate a set of clusters (Line 2). DBSCAN is a density-based clustering routine. It has the advantage that it can discover clusters of arbitrary shapes without specifying the number of clusters. Before the clustering process, we need to normalize the rating vectors in $L_{U,P}$ first. The normalization for each rating vector is achieved by having the value of each dimension divided by the sum of the value of each dimension. For example, suppose that a rating vector is $[0, 1, 4, 2, 1]$, then the normalized rating vector is $[0, 0.125, 0.5, 0.25, 0.125]$. The DBSCAN clustering approach works just like using a circle with the radius eps to scan the whole feature space from an arbitrarily selected point. The points will be grouped with the starting point together if they are

²We assume that the consumer has some ways to collect testimonies.

within the circle area of the starting point (we currently use 2-norm distance to calculate whether a point is within the circle area). Then the scanning process will continue by starting from the starting point and the points which are just included. The scanning process for this starting point will stop when no points are circle area reachable from all the points included in the starting point's cluster. All the points in the cluster are labeled as "touched". Then another arbitrary point which is not touched in last scan process is selected, and the same scanning process starts again. The whole clustering process will stop when no points are untouched. The most important parameter for the DBSCAN clustering approach to work correctly is the radius of the circle used to scan the feature space. Some work has been done to investigate the optimum radius value setting [69] [70]. In Section 4.5.1, we will carry out experiments to investigate the impact of the radius value on the filtering accuracy of iCLUB, and establish a feasible optimum radius value for iCLUB to work accurately. After the clustering process is completed, the Local component returns as honest witnesses the set of witnesses whose rating vectors are included in the same cluster as the consumer's rating vector (Lines 3-4). For simplicity, we use $R_{U,P}$ (or $R_{U_i,P}$) to denote U 's (or U_i) aggregated rating vector in the time window iCLUB is applied to.

4.3.2 Making Use of Global Information

As pointed out previously, the Local component is able to work effectively only when the consumer has some transactions with the target provider P . However, it is possible that the consumer does not have much experience with P in some scenarios, for example, where the consumer encounters the provider for the first time. In this case, we have to depend on the consumer's global information to filter unfair testimonies. That is the Global component. A pseudo code summary of how it works is given in Algorithm 3.

```

Procedure: Global( $P, U$ )
Input      :  $P$ , the provider whose reputation is evaluated;
                $U$ , the consumer evaluating  $P$ 's reputation;
Output     : A set of honest witnesses regarding  $P$ ;

1 foreach provider  $P_i$  ( $1 \leq i \leq M, P_i \neq P$ ) do
2   if  $U$  has transactions with  $P_i$  (i.e.,  $\|R_{U,P_i}\| \neq 0$ ) then
3      $W_i = \text{Local}(P_i, U)$ ;
4    $W_F = \bigcap_{i=1}^M W_i$ , where  $i$  satisfies  $\|R_{U,P_i}\| \neq 0$  and  $P_i \neq P$ ;
5    $C_1, C_2, \dots, C_L = \text{DBSCAN}(L_{U,P})$ ;
6   foreach cluster  $C_j$  ( $1 \leq j \leq L$ ) do
7      $W_{C_j} = \{U_i | R_{U_i,P} \in C_j\}$ ;
8     if  $W_F \neq \emptyset$  then
9        $W_{F_j} = W_F \cap W_{C_j}$ ;
10    else
11       $W_{F_j} = W_{C_j}$ ;
12   $q = \arg\{\max_j(|W_{F_j}|)\}, j = 1, 2, \dots, L$ ;
13 Return  $W_T = \{U_i | R_{U_i,P} \in C_q\}$  as honest witnesses;

```

Algorithm 3: Global component

The Global component first finds the honest witnesses for each provider with whom the consumer has transactions, using the Local() procedure (Lines 1-3). Then, a set of common honest witnesses W_F are formed as the intersection of the set of honest witnesses for each provider that the consumer has transactions with, except P (Line 4). The Global component continues by applying the DBSCAN routine to obtain the clustering result for P (Line 5). Then it calculates the intersection of W_F with the witnesses whose rating vectors are in each cluster achieved in Line 5 if W_F is not an empty set (Lines 6-11). Finally, it returns as honest witnesses the ones whose rating vectors are in the cluster which has the largest intersection result with W_F (Lines 12-13). Note that if two or more clusters are identified to contain the same largest size of intersection with W_F , the one that contains the consumer U 's rating vector (if any) will be considered as the cluster including honest witnesses.

In brief, the Global component of the iCLUB approach makes use of the consumer's experience with providers in the reputation system to find a set of witnesses who are honest regarding those providers, and then uses the information to find honest witnesses regarding the provider who is currently under evaluation. As noticed from Lines 4 and 12 in Algorithm 3, a witness may be considered as honest only if he has been honest regarding the providers encountered by the consumer. This restriction is based on the assumption that if a witness is honest for all the common providers encountered by the consumer, then it is more likely that the witness will be honest for the target provider. Another stronger restriction is that the witness has to belong to the largest intersection cluster regarding the target provider. This restriction is set to cope with collusion attacks where a group of witnesses collude in providing unfair testimonies to the target provider, by intentionally being honest regarding other providers to build up trust from the consumer. We will demonstrate this through experiments in Section 4.5.2.

4.3.3 The Integrated Approach

From the Global() procedure in Algorithm 3, we can see that the Global component has already integrated the Local() procedure (Lines 3-5). The iCLUB approach further integrates these two components using a threshold ε , as summarized in Algorithm 4.

Procedure: iCLUB(P, U)   Input : P, the provider whose reputation is evaluated; U, the consumer evaluating P's reputation;   Output : A set of honest witnesses regarding P; $\mathbf{if} \sum_{i=1}^K R_{U,P}(i) \geq \varepsilon \mathbf{then}$ 1 $\mid$ Return Local(P, U); 2 $\mid$ Return Local(P, U); 3 else 4 $\mid$ Return Global(P, U);

Algorithm 4: Integrate local and global components

There are some points worth of mentioning. First, the triggering of the use of global information is controlled by the threshold ε . If the number of transactions between the consumer U and P is smaller than ε , the Global component will be triggered. We will investigate how to properly set ε in Section 4.5.1. Second, it can be noticed that the Global component of iCLUB can work effectively when the consumer has no sufficient transactions with P but has transactions with other providers. However, for a consumer who is a newcomer to the system, he does not have any transaction with any provider. In this case, the Global component has to follow the majority rule as other approaches do, such as the iterated filtering approach [1] and the clustering filtering approach [45].

4.4 Examples

In this section, we use some examples to demonstrate how the iCLUB approach works. These examples involve 10 witnesses, indexed by $W_1, W_2, \dots, W_{10}$, 1 consumer U and 5 providers, indexed by $P_1, P_2, \dots, P_5$. Five rating levels are used as an illustration. The radius used for DBSCAN clustering is set as $eps = 0.3$, and the threshold to trigger the global component is set as $\varepsilon = 10$.

4.4.1 Local Component Example

When there are more than ε transactions between the consumer and the target provider, the Local component is called. Suppose that P_5 's reputation is under evaluation. W_1 to W_8 are dishonest witnesses. The numbers of transactions rated as rating level 1 to 5 by each witness and U for the 5 providers are shown in Table 4.1.

After DBSCAN clustering, there are two clusters achieved. $W_1, W_2, W_3, W_4, W_5, W_6, W_7$, and W_8 are in one cluster. W_9, W_{10} and U are in the other cluster. As indicated in

	Reported Rating Vectors					Normalized Rating Vectors				
Rating Level	1	2	3	4	5	1	2	3	4	5
W_1	0	0	0	0	31	0	0	0	0	1.0000
W_2	0	0	0	0	14	0	0	0	0	1.0000
W_3	0	0	0	0	21	0	0	0	0	1.0000
W_4	0	0	4	6	33	0	0	0.0930	0.1395	0.7674
W_5	0	0	2	4	6	0	0	0.1667	0.3333	0.5000
W_6	0	0	2	14	32	0	0	0.0417	0.2917	0.6667
W_7	0	0	4	10	26	0	0	0.1000	0.2500	0.6500
W_8	0	0	0	0	30	0	0	0	0	1.0000
W_9	18	12	4	0	0	0.5294	0.3529	0.1176	0	0
W_{10}	12	6	2	0	0	0.6000	0.3000	0.1000	0	0
U	17	8	4	0	0	0.5862	0.2759	0.1379	0	0

Table 4.1: Rating vectors for the local component working example

Algorithm 2, the witnesses whose rating vectors are in the cluster including the consumer's rating vector are returned as honest witnesses. W_9 and W_{10} are considered as honest witnesses. This result is consistent with the initial setting.

4.4.2 Global Component Example

When the consumer does not have enough transactions with the target provider (in this example, it is when the consumer has less than 10 transactions with P_5), the Global component is triggered. Suppose that whether the witnesses are honest for the 5 providers are shown in Table 4.2.

	Dishonest	Honest
S_1	W_1	$W_2, W_3, W_4, W_5, W_6, W_7, W_8, W_9, W_{10}$
S_2	W_1, W_2, W_3	$W_4, W_5, W_6, W_7, W_8, W_9, W_{10}$
S_3	W_2, W_4	$W_1, W_3, W_5, W_6, W_7, W_8, W_9, W_{10}$
S_4	$W_1, W_2, W_3, W_4, W_5, W_6, W_7$	W_8, W_9, W_{10}
S_5	W_5, W_8	$W_1, W_2, W_3, W_4, W_6, W_7, W_9, W_{10}$

Table 4.2: The witnesses' honesty regarding respective providers

Also suppose that the consumer does not have enough transactions with P_5 whose

reputation is under evaluation, but he has enough transactions with other 4 providers (P_1, P_2, P_3 , and P_4). We first use DBSCAN to cluster the testimonies for the 4 providers. By assuming the clustering result is consistent with the setting, we can get W_F :

$$\begin{aligned}
W_F &= \{W_2, W_3, W_4, W_5, W_6, W_7, W_8, W_9, W_{10}\} \\
&\bigcap \{W_4, W_5, W_6, W_7, W_8, W_9, W_{10}\} \\
&\bigcap \{W_1, W_3, W_5, W_6, W_7, W_8, W_9, W_{10}\} \\
&\bigcap \{W_8, W_9, W_{10}\} \\
&= \{W_8, W_9, W_{10}\}.
\end{aligned} \tag{4.4}$$

Second, we use DBSCAN to cluster P_5 's testimonies, and assume that we get two clusters – C_1 including W_5 and W_8 's testimonies, and C_2 including $W_1, W_2, W_3, W_4, W_6, W_7, W_9$ and W_{10} 's testimonies:

$$W_{C_1} = \{W_5, W_8\}, \tag{4.5}$$

$$W_{C_2} = \{W_1, W_2, W_3, W_4, W_6, W_7, W_9, W_{10}\}. \tag{4.6}$$

Then we calculate the intersection of W_F with W_{C_1} and W_{C_2} to get W_{F_1} and W_{F_2} , respectively:

$$W_{F_1} = W_F \bigcap W_{C_1} = \{W_8\}, \tag{4.7}$$

$$W_{F_2} = W_F \bigcap W_{C_2} = \{W_9, W_{10}\}. \tag{4.8}$$

As the cluster C_2 has a larger size of intersection set with W_F , $W_1, W_2, W_3, W_4, W_6, W_7, W_9$ and W_{10} whose rating vectors are in C_2 are considered as honest witnesses for P_5 , which is consistent with the initial setting.

4.5 Experimental Studies

We carry out four sets of experiments to evaluate the proposed iCLUB approach. The first set investigates the relationship between the radius value of DBSCAN and the accuracy of iCLUB in filtering unfair testimonies, and explores the appropriate ε value used to trigger the Global component. The aim of the second set of experiments is to examine the accuracy of iCLUB in various scenarios, and in particular, its robustness against collusion attacks. The third set of experiments is to integrate the iCLUB approach with DRS [59] to examine whether the iCLUB approach will improve the reputation system's robustness. The fourth set is to compare iCLUB with other representative approaches (the iterated filtering approach [1] and TRAVOS [48]) in terms of the accuracy of filtering unfair testimonies and evaluating provider reputation.

In our experiments, we still use FPR (Equation (3.7)), FNR (Equation (3.8)) and MCC (Equation (3.9)) to measure the accuracy of filtering unfair testimonies. In our experiments investigating the accuracy of filtering unfair testimonies when multi-nominal rating levels are applied, four types of dishonest witnesses are involved:

1. *ballot-stuffing* witnesses [45] give testimonies that providers behave well regardless of the true behaviors of the providers;
2. *badmouthing* witnesses [45] give testimonies that providers behave bad regardless of the true behaviors of the providers.
3. *γ -low-shifting* witnesses give testimonies that are γ levels lower than the real ratings.
4. *γ -high-shifting* witnesses give testimonies that are γ levels higher than the real ratings.

4.5.1 Choosing Proper Parameters

This set of experiments is to investigate how the radius value of DBSCAN clustering will impact the accuracy of iCLUB in filtering unfair testimonies and how to set ε value in Algorithm 4 to trigger the Global component. To investigate the influence of DBSCAN radius value, we focus on the Local component as it is sufficient to explore the impact. We simulate a service provision environment that involves 1 provider P , ω witnesses, and 1 consumer U . We simulate 10,000 rounds for each simulation scenario to achieve a statistical accuracy. In each round of a simulation, an initial willingness (iw) value ($0 \leq iw \leq 1$) is randomly generated for the provider to represent how much the provider is willing to cooperate. According to the generated iw value, different types of dishonest witnesses are generated for different iw value ranges. The relationship between iw value and the types of dishonest witnesses is shown in Table 4.3.

iw value	Types of Dishonest Witnesses
$0 \leq iw \leq 0.3$	ballot-stuffing, γ -high-shifting
$0.3 < iw < 0.7$	badmouthing, ballot-stuffing, γ -low-shifting, γ -high-shifting
$0.7 \leq iw \leq 1$	badmouthing, γ -low-shifting

Table 4.3: Relationship between iw value and types of dishonest witnesses

Each witness or U has T transactions with the provider P . For each transaction, one willingness value will be generated from a normal distribution whose mean is iw , and standard deviation is δ . The mapping between the willingness value for each transaction and the rating level for P is shown in Table 3.1. In this way, the provider's behavior is represented by the normal distribution corresponding to his initial willingness value. For example, if P 's initial willingness value is 0.2, most of the ratings for him should be 1. When honest witnesses or the consumer has more transactions with P , his behavior will be represented by their rating vectors more precisely.

We explore how the radius value will impact the accuracy of iCLUB for different

levels of scalability and stability. Here scalability represents the number of witnesses who will report ratings to the consumer, and stability represents the number of transactions between P and the witnesses (or U). A high scalability means that there are a lot of witnesses. On the contrary, a low scalability means that there are only a few witnesses. A high stability means that there are a large number of transactions between P and the witnesses (or U), and a low stability means that there are only a few transactions between P and the witnesses (or U). We simulate two levels of scalability and stability – high (i.e., 100 witnesses and 100 transactions) and low (i.e., 10 witnesses and 10 transactions).

Table 4.4 lists the parameter meanings and values we use in our simulation.

Parameter	Meaning	Value
ω	The number of witnesses	$\{10, 100\}$
T	The number of transactions between each witness or U and P	$\{10, 100\}$
δ	The standard deviation of the normal distribution to simulate P 's behavior	$\{0.2, 0.3\}$
P_{unfair}	The percentage of the dishonest witnesses	$\{40\%, 80\%\}$
γ	The shifting level of the reported ratings from the γ -low-shifting witnesses or the γ -high-shifting witnesses	$\{1, 2, 3, 4\}$

Table 4.4: Simulation parameters, meanings and values

In this table, P_{unfair} is the total percentage of the dishonest witnesses. In each round of a simulation scenario, the percentage of each type of dishonest witnesses is equally distributed. For example, if $p_{unfair} = 40\%$, and there are badmouthing witnesses and γ -low-shifting witnesses, the percentage of each type of dishonest witnesses is 20%. In general, we have 64 parameter value combinations. Therefore, there are 64 scenarios simulated in total. As an illustration, we show the results of the MCC, FPR, and FNR value changes with DBSCAN radius (i.e., ϵ) increasing from 0.1 to 1.4 when $\gamma = 2$ and $\delta = 0.2$.

Figure 4.1 shows the MCC, FPR and FRR results when the scalability is low (i.e.,

$\omega = 10$) and the stability is also low (i.e., $T = 10$). According to the results, a higher MCC value can be achieved when the DBSCAN radius value is in the range of $[0.3, 0.5]$. When the radius value is too small (i.e., 0.1 to 0.3), there are more false negatives, meaning that there are more honest witnesses misclassified as dishonest witnesses. When the radius value increases from 0.5 to 1.4, there are more false positives, meaning that there are more dishonest witnesses misclassified as honest witnesses. The reason is as follows. When the stability is low, some honest witnesses or the consumer's rating vector may not represent the provider's behavior. When the radius is small, these honest witnesses cannot be grouped into the same cluster as the consumer, resulting in that they are incorrectly filtered as dishonest witnesses. When the radius is large, such as 1.4, the dishonest witnesses are included in the honest witnesses' cluster as the DBSCAN scanning radius is too large to differentiate the honest witnesses' rating vectors from the dishonest witnesses' rating vectors. Therefore, the FPR value increases with the radius value.

Figure 4.1: Varying radius when $T=10$, $\omega = 10$, and $P_{unfair} = 40\%$

Figure 4.2 shows the MCC, FPR and FNR results when the scalability is low (i.e., $\omega = 10$) and the stability is high (i.e., $T = 100$). As there are more transactions between each witness or the consumer and the provider, the honest witnesses or the consumer's rating vector can represent the provider's behavior more accurately. Compared to the results when the stability is low, there is a larger workable radius range, which is about $[0.2, 0.6]$. Similar to the results when the stability is low, a smaller radius value will lead

to a larger FNR value, and a larger radius value will lead to a larger FPR value.

Figure 4.2: Varying radius when $T=100$, $\omega = 10$, and $P_{unfair} = 40\%$

Figure 4.3 shows the MCC, FPR and FNR results when the scalability is high (i.e., $\omega = 100$) and the stability is low (i.e., $T = 10$). It can be noticed that the workable radius value range is quite small. As when the stability is low and there are a lot of witnesses, there is a larger diversity among the witnesses' rating vectors, which makes the correct clustering result difficult to be achieved. Similar to the results in the previous two scenarios, more errors happen as false positives when the radius value increases.

Figure 4.3: Varying radius when $T=10$, $\omega = 100$, and $P_{unfair} = 40\%$

Figure 4.4 shows the MCC, FPR and FNR results when the scalability is high (i.e., $\omega = 100$) and the stability is also high (i.e., $T = 100$). Similar to the results shown in Figure 4.2, there is a larger workable radius range. It can also be noticed that there are almost no false negatives. As when the stability is high, the honest witnesses or the

consumer's rating vector can reflect the provider's behavior more precisely. A small radius value (i.e., $[0.1, 0.3]$) can also correctly differentiate the honest witnesses from dishonest witnesses.

Figure 4.4: Varying radius when $T=100$, $\omega = 100$, and $P_{unfair} = 40\%$

As an illustration, Figure 4.5 shows the MCC results when $P_{unfair} = 80\%$ for the four scenarios (the scalability is high or low, and the stability is high or low). We can see that MCC presents the similar trend as the results of the corresponding scenarios when $P_{unfair} = 40\%$.

According to the simulation results (including the results not presented here), it is difficult to find an appropriate radius value when $\gamma = 1$ as there is only a small difference between γ -shifting witnesses' rating vectors and the honest witnesses' rating vectors. But from another point of view, as the difference is quite small, we can consider it as a subtle subjective difference and treat these witnesses as honest witnesses. When $\gamma = 2, 3$ or 4 , a workable radius value range can be found. Generally speaking, the range is larger when the stability is higher, and it is smaller when the stability is lower. A very small radius value (i.e., 0.1) will lead to a larger FNR value. And a larger radius value will lead to a larger FPR value. Therefore, if FNR is more concerned, a too small radius value should not be adopted. If FPR is more concerned, then a small or medium radius value can be applied. As we have mentioned earlier, we currently do not focus on how to collect testimonies. But we suggest here that when collecting testimonies and evaluating the

Figure 4.5: Varying radius when $P_{unfair} = 80\%$

provider's reputation, the witnesses who have a low stability should not be considered as eligible since their testimonies may not reflect their behaviors or the provider's behavior, which will lead to an incorrect filtering result, especially when the scalability is high. By summarizing the simulation results, the workable radius value more falls into the range of $[0.3, 0.5]$. As we are more concerned with the FPR value, we use 0.3 as the DBSCAN radius value in the following simulations. The δ value is randomly selected (either 0.2 or 0.3) in each round of the following simulations.

To investigate how to properly set ε value which is used to trigger the Global component, we simulate three scenarios. In the first scenario, the frequency of the witnesses having transactions with P is lower than that of U . In the second scenario, the frequency of the witnesses having transactions with P is the same as that of U . In the third scenario, the frequency of the witnesses having transactions with P is higher than that of

U . Here the frequency represents how often a witness or U will have one transaction with P . If we use U 's frequency as the baseline, the equal frequency scenario means that when U has one transaction with P , each witness also has one transaction with P . The lower frequency scenario means that when each witness has one transaction with P , U already has multiple transactions with P . The higher frequency means that when U has one transaction with P , each witness already has multiple transactions.

Figure 4.6 shows the changes of MCC values with the number of transactions increasing in half frequency, equal frequency, and double frequency scenarios, respectively, where half frequency means that when U has 2 transactions with P , a witness has only 1 transaction with P from the view of probability.

Figure 4.6: (a) Frequency of consumer transaction is double as that of witnesses; (b) Equal frequency of transacting with the provider; (c) Frequency of witness transaction is double as that of consumer.

According to the results, in the half frequency scenario, the Local component achieves a stable filtering result after U having about 25 transactions. In the equal frequency scenario, the Local component achieves a stable filtering result after U having about 12 transactions. In the double frequency scenario, the Local component achieves a stable filtering result after U having about 8 transactions. These results suggest that the setting of ε value should be related to the relative frequency of the witnesses and U having transactions with the provider. It is a good practice for U to evaluate the frequency of

the transactions between the witnesses and the provider before he sets the ε value. If the frequency of witnesses having transactions is higher than or equal to that of the consumer, then a small ε value can be adopted. When the frequency of the witnesses is lower than that of the consumer, a larger ε value is more preferred. In our following experiments, we use $\varepsilon = 10$ because the equal frequency scenario is simulated.

4.5.2 Robustness against Collusion Attacks

The goal of this set of experiments is to investigate the accuracy of the iCLUB approach in filtering unfair testimonies. In particular, we investigate the robustness of iCLUB against collusion attacks (i.e., some witnesses collude together to give unfair testimonies for some particular providers). As demonstrated in the first set of experiments, the Local component of the iCLUB approach can work well when the consumer has some transactions with the target provider. But it is obvious that when the consumer has no sufficient number of transactions with the target provider, we need the Global component to facilitate the filtering of unfair testimonies.

In this experiment, we simulate a more complicated service provision environment that involves 10 providers, 100 witnesses and 1 consumer U . Each provider is attached with a profile, describing his iw value range and the percentage of dishonest witnesses for him. The first 200 transactions of each witness or U are for the presetting stage. In this stage, the witnesses will randomly select one provider among the 10 providers as the partner for each transaction, and U will randomly select one provider among the first 9 as the partner for each transaction, leaving the last provider alone to investigate the accuracy of iCLUB (especially the Global component) in filtering unfair testimonies. After the presetting stage, U will randomly select one provider among the 10 providers as partner for each transaction. We simulate two kinds of scenarios: the witnesses' behaviors keep

consistent for all the providers, and the witnesses' behaviors change over providers.

When the witnesses' behaviors keep consistent, the dishonest witnesses will report ratings unfairly for all providers and the honest witness will report ratings for all providers in an honest way. In each round of this simulation, the percentage of dishonest witnesses increases from 10% to 90% for each provider. The providers' iw values are randomly generated and the percentage of each type of dishonest witnesses is also randomly generated according to the provider's iw value (the sum of the percentage of each type of dishonest witnesses should be equal to P_{unfair}). Figure 4.7(a) shows the tenth provider's MCC value changes with the percentage of dishonest witnesses increasing from 10% to 90%. It can be seen that the Global component can work well no matter what the percentage of the dishonest witnesses is.

Figure 4.7: (a) The witnesses' behaviors keep consistent for all providers; (b) The honest witnesses' behaviors keep consistent for all providers; (c) The witnesses' behaviors change over providers; (d) MCC varying with the number of transactions when $P_{unfair} = 60\%$

We then investigate the accuracy of the Global component working when the witnesses' behaviors change. First, we simulate a specific scenario – some witnesses keep honest for all the providers. To simulate this scenario, we set the profiles of the 10 providers as shown in Table 4.5.

Provider Index	P_1	P_2	P_3	P_4	P_5
iw	$[0, 0.3]$	$(0.3, 0.7)$	$[0.7, 1]$	$[0, 0.3]$	$[0, 0.3]$
P_{unfair}	0	0	0	20%	60%
Provider Index	P_6	P_7	P_8	P_9	S_{10}
iw	$[0.7, 1]$	$[0.7, 1]$	$(0.3, 0.7)$	$(0.3, 0.7)$	varying
P_{unfair}	20%	60%	30%	90%	varying

Table 4.5: Profiles of the providers

P_1 , P_2 and P_3 are used to simulate the providers who have no dishonest witnesses. P_4 and P_5 are used to simulate the providers who have ballot-stuffing witnesses and γ -high-shifting witnesses. They represent two scenarios – one is that the percentage of the dishonest witnesses is smaller than that of the honest witnesses, and the other one is that the percentage of the dishonest witnesses is larger than that of the honest witnesses. P_6 and P_7 are used to simulate the providers who have badmouthing witnesses and γ -low-shifting witnesses. They also represent the two scenarios as what P_4 and P_5 do. P_8 and P_9 are used to simulate the providers who have four types of dishonest witnesses at the same time. They also represent the two scenarios as what P_4 and P_5 do. The initial willingness value and percentage of dishonest witnesses for P_{10} will change in the simulation to investigate the accuracy of the iCLUB approach in varying scenarios where, for example, the provider may change his behavior over time. For each provider, the first P_{unfair} percentage of witnesses are generated as dishonest witnesses, and other witnesses are generated as honest witnesses. Through this setting, only the last 10% witnesses are honest for all providers. Though 61% to 90% witnesses are honest for the first eight providers, they are detected as dishonest for provider P_9 . If they collude again for provider P_{10} , they still can be detected as dishonest.

Figure 4.7(b) shows P_{10} 's MCC changes with dishonest witness percentage increasing from 10% to 90%. According to the results, when a certain number of witnesses keep honest for all the providers, the Global component can work well. The reason behind is as follows. As described in Algorithm 3, the Global component takes the intersection of all the honest witnesses for all the providers encountered by U as W_F . If a witness keeps honest for all the providers, then he must be in W_F (assuming that the clustering result is correct). For the target provider P_{10} , suppose there are five clusters achieved after clustering – C_1 including the badmouthing witnesses' testimonies, C_2 including the ballot-stuffing witnesses' testimonies, C_3 including the γ -low-shifting witnesses' testimonies, C_4 including the γ -high-shifting witnesses' testimonies, and C_5 including the honest witnesses' testimonies. As the witnesses in W_F are honest for all the providers (including P_{10}), their testimonies should not be in C_1 , C_2 , C_3 , or C_4 . Then C_5 will have the largest intersection result with W_F . Therefore, the Global component can get the correct filtering result.

Keeping the second scenario that the witnesses change behaviors, we remove the assumption that some witnesses keep honest for all the providers by simulating the scenario where the witnesses' behaviors change over providers randomly. In this simulation, the 10 providers' iw value ranges are kept the same as in last simulation. The percentage of the unfair witnesses is the same for the 10 providers. For each provider, the dishonest witnesses are randomly assigned. Figure 4.7(c) shows P_{10} 's MCC value changes with P_{unfair} increasing from 10% to 90%. According to the results, the Global component can work well when P_{unfair} is smaller than 30%. When P_{unfair} is larger than 50%, the filtering result is almost like random guessing.

Though the Global component cannot work when P_{unfair} is larger than 30% and the witnesses' behaviors change randomly over providers, the problem can still be solved if U can sacrifice some transactions with the target provider. Figure 4.7(d) shows the changes of MCC value for P_{10} with the number of transactions increasing after the presetting

stage when we set $\varepsilon=10$. The transaction number starts from 0 and ends at 150. Though for a particular transaction, the consumer may not select P_{10} as his partner, we still calculate the MCC value according to the iCLUB filtering result for P_{10} . The percentage of dishonest witnesses is set as 60%. It can be noticed that after about 120 transactions, the MCC value approaches to 1. Since a provider is randomly selected among 10 providers as the partner for each transaction, the consumer may only sacrifice 12 transactions to accumulate the experiences for the iCLUB approach to cope with the situation where 60% witnesses randomly collude. Actually, this number is quite affirmative with the ε value to trigger the global component we get in the experiment of exploring the ε value for the equal frequency scenario in Section 4.5.1.

4.5.3 Integration with Dirichlet Reputation System

The iCLUB approach is a filtering approach to address the problem of unfair testimonies instead of a reputation system. But the iCLUB approach can be integrated with reputation systems where consumers' rating vectors are available. In this part, we use the Dirichlet Reputation System (DRS) [59] as an example to show that the iCLUB approach can contribute to improving the reliability of a reputation system.

DRS works as follows. Let the aggregated rating vector with aging from consumer U_n regarding provider P_m be A_{U_n, P_m}^f until time window f . Then the aggregated rating vector regarding P_m until time window f , $A_{P_m}^f$, is:

$$A_{P_m}^f = \sum_{n=1}^N A_{U_n, P_m}^f, \quad (4.9)$$

DRS assumes that a provider's behavior follows a Dirichlet probability distribution [40]. When mapping to K rating levels, the expected probability p_k that P_m will behave as

rating level k ($1 \leq k \leq K$) in the future is:

$$E(p_k) = \frac{A_{S_m}^f(k) + \frac{C}{K}}{C + \sum_{k=1}^K A_{S_m}^f(k)}, \quad (4.10)$$

where C is a *a priori* constant [40] (“ C will be always equal to the cardinality of the state space over which a uniform distribution is assumed. For example, $C = 2$ when a uniform distribution over a binary state space is assumed” [59]). As a further step, we evaluate P_m ’s reputation $\hat{E}_{P_m}$ as follows:

$$\hat{E}_{P_m} = \sum_{k=1}^K E(p_k) \times K. \quad (4.11)$$

We simulate a similar service provision environment as that in Section 4.5.2, which includes 1 consumer, 10 providers and 100 witnesses. Ten time windows are simulated. In each time window, the providers’ behaviors do not change. After each time window, the providers’ behaviors will change. The providers’ behavior changes are simulated by generating different *iw* value ranges. For example, if in the first time window, the provider P_1 ’s *iw* value falls into the range of $[0,0.3]$, then in the second time window, his *iw* value will fall into the range of $(0.3,0.7)$ or $[0.7,1]$. In each time window, the number of the transactions from each witness or the consumer is a randomly generated number in the range of $[100,200]$. The unfair percentage for each provider is 0 for P_1 , 10% for $P_2, \dots$, and 90% for P_{10} . The percentage of each type of dishonest witnesses for each provider is randomly generated. In each time window, iCLUB is applied to get the honest witnesses, and only the achieved honest witnesses’ ratings are passed to DRS to calculate the providers’ reputation. $C = 5$ is used as DRS *a priori* constant. We measure the accuracy of evaluating provider reputation using the Mean Absolute Error (MAE) as

follows:

$$MAE = \frac{\sum_{m=1}^M |E_{P_m} - \hat{E}_{P_m}|}{M}, \quad (4.12)$$

where M is the number of providers, E_{P_m} is the provider P_m 's expected reputation which is calculated using the honest witnesses' ratings, and $\hat{E}_{P_m}$ is P_m 's evaluated reputation using the achieved honest witnesses' ratings after applying iCLUB. Figure 4.8 shows the MAE changes with forgetting factor λ and the number of windows. The result shows that MAE is quite small after applying iCLUB. It implies that the evaluated provider reputation after using the iCLUB approach is very close to the expected provider reputation.

Figure 4.8: MAE changes with forgetting factor and the number of windows

4.5.4 Comparative Experiments

In this experiment, we compare the iCLUB approach with other approaches from two aspects – the accuracy of filtering unfair testimonies and evaluating provider reputation. We compare iCLUB with the iterated filtering approach [1] in terms of filtering accuracy. We also compare iCLUB with both the iterated filtering approach and TRAVOS [48] in terms of the accuracy of evaluating provider reputation. Note that in the comparative experiments, after filtering unfair testimonies using iCLUB or the iterated filtering approach, fair testimonies will be aggregated to evaluate provider reputation in the same way as that used in BRS [39].

A similar service provision environment as that in the second set of experiments is simulated. The difference is the way to generate the rating for each transaction as there are only two rating levels in this experiment. The initial willingness (iw) value assigned to each provider is taken from the value set $\{0.1, 0.2, 0.4, 0.6, 0.8, 0.9\}$. To simulate provider behavior changes, a provider's first transaction uses the initial willingness value. In the following transactions, a willingness value is generated through one of the three strategies – the willingness value of last transaction subtracting 0.02, equal to the willingness value of last transaction, or the willingness value of last transaction adding 0.02. The three strategies are uniformly selected during the simulation. The willingness value for each transaction is also limited in the range of $[iw-0.1, iw+0.1]$. The willingness value indicates in what probability the provider's behavior will be rated as positive. Compared to the experiments for multi-nominal ratings, there are no γ -low-shifting or γ -high-shifting witnesses. For the case of binary ratings, we study three types of dishonest witness: 1) ballot-stuffing witnesses; 2) badmouthing witnesses; 3) opposite witnesses who report the opposite of the real ratings as testimonies. We assume that a provider with a larger iw value (i.e., 0.8 and 0.9) will not have ballot-stuffing witnesses, a provider with a smaller

iw value (i.e. 0.1 and 0.2) will not have badmouthing witnesses and a provider with a medium iw value (i.e. 0.4 and 0.6) will not have opposite witnesses. Table 4.6 shows the profiles of the 10 providers.

Provider Index	P_1	P_2	P_3	P_4	P_5
iw	0.1	0.4	0.8	0.1	0.2
P_{unfair}	0	0	0	40%	80%
Provider Index	P_6	P_7	P_8	P_9	P_{10}
iw	0.8	0.9	0.4	0.6	varying
P_{unfair}	40%	80%	40%	80%	varying

Table 4.6: Profiles of the providers

We keep the first 200 transactions as the presetting stage. Each witness will randomly select one provider among the 10 providers as his partner for each transaction, and the consumer U will randomly select one provider among the first 9 as his partner. Figure 4.9 shows P_{10} 's MCC value changes for iCLUB and the iterated filtering approach with the percentage of dishonest witnesses increasing from 10% to 90% after the presetting stage. It can be noticed that the accuracy of the iterated filtering approach decreases with the increase of the percentage of dishonest witnesses. The iCLUB approach performs stably until a significant percentage ($> 85\%$) of witnesses are dishonest.

Figure 4.10 shows P_{10} 's MCC value changes with the increase of the number of transactions after the presetting stage when the percentage of dishonest witnesses is 90%. We set $\varepsilon=10$. It can be noticed that after about 100-150 transactions, which is about 10-15 transactions between U and P_{10} , the MCC value using iCLUB is close to 1.

Figure 4.11 shows the comparison result of reputation evaluation for P_{10} with the percentage of dishonest witnesses by using the iterated filtering approach, iCLUB and TRVOS after the presetting stage. The expected reputation value is calculated using the honest witnesses' testimonies only. And the unfair reputation value is calculated using all witnesses' testimonies.

Figure 4.9: MCC changes with the percentage of dishonest witnesses

Figure 4.10: MCC changes with the number of transactions

Figure 4.11: Reputation evaluation value changes with the percentage of dishonest witnesses

According to the results, when $iw = 0.1$, $iw = 0.2$, $iw = 0.8$, and $iw = 0.9$, the reputation value after using iCLUB is initially close to the expected reputation value. With the increase of the percentage of dishonest witnesses, the reputation value after using iCLUB has a small deviation from the expected reputation value until reaching a significant percentage of dishonest witnesses ($> 85\%$). The reputation value after using the iterated filtering approach is initially close to the expected reputation value, and then continuously deviates from the expected reputation value. After the percentage of the dishonest witnesses is greater than 50%, the reputation value after using the iterated filtering approach is even worse than the unfair reputation value. The reputation value after using TRAVOS is initially worse than that using the iterated filtering approach or iCLUB, and when the percentage of dishonest witnesses is greater than 40%, it is better than that using the iterated filtering approach, but still worse than that using iCLUB. When $iw = 0.4$ and $iw = 0.6$, the expected reputation value, the unfair reputation value

and the reputation value after using the iterated filtering approach, TRAVOS or iCLUB are quite close to one another since there exist badmouthing witnesses and ballot-stuffing witnesses at the same time. The impacts of the two types of dishonest witnesses counteract each other.

Figure 4.12 shows the comparison result for P_{10} 's reputation evaluation with the number of transactions increasing after the presetting stage when the percentage of dishonest witnesses is 90%. ε is set as 10. We can see that after about 10-15 transactions (100-150 on the x-axis in the figure), the evaluated reputation value using iCLUB is very close to the expected reputation value, indicating that iCLUB achieves a more accurate filtering result than the iterated filtering approach and TRAVOS.

Figure 4.12: Reputation evaluation value changes with the number of transactions

4.5.5 Discussions on Experimental Results

According to the experimental results, the iCLUB approach presents a promising filtering accuracy when being applied to both reputation systems with multi-nominal rating levels and those with binary ratings. To make iCLUB work more effectively, the DBSCAN radius value is very important. Though we have established a feasible DBSCAN radius value range for iCLUB through experimentation, it is worth pointing out that this value range is dependent on the simulation data. We argue that a smaller radius value may be better though it may cause more false negatives but it can decrease the number of false positives. As our aim is to filter unfair testimonies, false positives should carry more weight. Another suggestion is that it is better not to consider the testimonies from the witnesses who have few transactions with the providers when collecting testimonies. As such testimonies may not reflect the witnesses' or the providers' behaviors, these unstable testimonies may cause a wrong clustering result.

The experimental results shown in Section 4.5.1 suggest that there is a relationship between the ε value setting and the frequency of consumer having transactions with providers. But in the scenario where a large number of witnesses collude randomly, ε should also be set smaller to disable the Global component at an earlier stage as the Global component cannot address this extreme situation. According to the experimental results, setting $\varepsilon=10$ is a good choice in many cases though iCLUB has not achieved its best performance in the low frequency scenario (as shown in Figure 4.6(a)).

The accuracy of the Global component is dependent on two factors – the Local component working accuracy and the witnesses' behaviors. The Local component working accuracy can be optimized by adopting an appropriate clustering parameter value. But the witnesses' behaviors are not controlled by iCLUB. Therefore, in the scenario where the witnesses' behaviors keep consistent for the providers or at least the honest witnesses'

behaviors keep consistent for the providers, the Global component can work accurately. But in the scenario where the witnesses' behaviors keep changing over the providers, the consumer has to sacrifice some transactions with the target provider to get an accurate filtering result. The accuracy of the Global component is also dependent on the providers from whom the global information is obtained. If it happens that the Global component only collects the testimonies for the providers to whom the witnesses are honest and does not collect the testimonies for the providers to whom the witnesses are dishonest, then a wrong filtering result may also be led to.

In the comparative experiments, that the iCLUB approach is better than the iterated filtering approach is due to two reasons. First, iCLUB uses global information to filter the unfair testimonies. Second, iCLUB assigns the consumer's personal ratings with a higher weight (the Local component will keep as fair the testimonies from the witnesses whose rating vectors are in cluster including the consumer's rating vector). The TRAVOS approach also uses global information and assigns the consumer's personal ratings with a higher weight, but TRAVOS needs more transactions to achieve a high accuracy in evaluating provider reputation compared to the iCLUB approach.

4.6 Summary

In this chapter, we proposed the iCLUB approach to filter unfair testimonies to build reliable reputation systems. The iCLUB approach preserves the merits that the two-stage clustering filtering approach has. For example, it can resist the flooding attack by normalizing the collected testimonies. Besides the preserved merits, the iCLUB approach also improves the two-stage clustering approach in the following aspects:

- By considering global information, it can work well when the consumer's personal ratings regarding the target provider is unavailable and majority witnesses are dis-

honest. As the results shown in Section 4.5.2, iCLUB can achieve a high filtering accuracy when the consumer has no personal ratings regarding the provider and even 90% witnesses are dishonest.

- It is able to cope with collusion attacks to a great extent. The results in Section 4.5.2 show that iCLUB can achieve a high filtering accuracy when 90% witnesses consistently collude together and up to 30% witnesses randomly collude together. Even if in the scenario where more than 30% witnesses collude together, iCLUB can resist the collusion attacks by using Local component.

But it is worth pointing out that iCLUB still cannot work well when the consumer is a newcomer to the system and unfortunately majority witnesses are dishonest for the target provider. In such scenario, we have to depend on other possible information to address the problem of unfair testimonies. Such possible information can be the relationship between the consumer and the witnesses, the relationship between the witnesses and the provider, and the frequency of the providers receiving ratings [27] [71] [72].

Chapter 5

A Witness Trustworthiness Model

We presented in Chapter 3 a two-stage clustering filtering approach to address the problem of unfair testimonies using personal, public, and local information. To improve the two-stage clustering filtering approach, we proposed in Chapter 4 an integrated clustering filtering approach by further incorporating global information. A major disadvantage of the two approaches is that they depend on an optimized clustering threshold setting to achieve an accurate filtering result. This may require a large amount of data for the training purpose. To avoid this problem, we propose in this chapter a witness trustworthiness model¹ based on Dempster-Shafer theory [50] [51] to calculate the witness trustworthiness using the modeling mechanism as the approach categorization proposed in Chapter 2.

5.1 Introduction

In Chapter 3 and 4, two clustering-based filtering approaches were proposed to address the problem of unfair testimonies for reputation systems. Though as experimental results show, the two approaches can effectively mitigate the influence of unfair testimonies and contribute to building reliable reputation systems, they require an optimized parameter

---

¹This work is partially published in ICEC2012 (conference paper [2] in author's publication list) and AAMAS2012 (conference paper [3] in author's publication list).

setting to achieve an accurate filtering result. This may not be easily achieved in reality.

To address this limitation, we propose in this chapter a novel *WITness Trustworthiness* model (WIT) based on Dempster-Shafer theory. More specifically, WIT incorporates personal, public, local and global information to model the trustworthiness for a particular witness. The different pieces of information are combined together by using Dempster-Shafer theory to obtain the witness trustworthiness value, which is specific for different providers. Moreover, we use the achieved witness trustworthiness value as a discounting factor to discount the testimonies reported by the witness. Then the provider's reputation is evaluated by aggregating the witnesses' discounted testimonies. A service provision environment is simulated to examine the performance of WIT. Experimental results show that WIT can model the witness trustworthiness accurately and flexibly. Compared with the approaches reviewed in Chapter 2 that considers global information [47] [48] [2] [60] [64], WIT does not depend on the assumption that the witness behaves consistently over all providers. Therefore, WIT has the capability to work in the scenario where the witness behavior changes over providers.

The remainder of this chapter is structured as follows. Section 5.2 introduces the Dempster-Shafer theory which is the foundation of WIT. Section 5.3 presents WIT, including the witness trustworthiness model and witness rating discounting scheme. Section 5.4 presents the experimental studies regarding the performance of WIT. Some issues requiring more consideration are discussed in Section 5.5. Finally, Section 5.6 summarizes this chapter.

5.2 Dempster-Shafer Theory Framework

Before coming into the details of WIT, we first briefly introduce the Dempster-Shafer theory [50] [51]. We will use a binary case to show some basic definitions in Dempster-

Shafer theory and how the theory works.

Let a and b be two ratings a provider will be rated as (e.g., a = positive, b = negative). For the frame of discernment $\Theta = \{a, b\}$, we have the power set $2^\Theta = \{\{a\}, \{b\}, \{a, b\}, \emptyset\}$ which is the set of propositions regarding the actual rating the provider is rated as. Here the proposition that $\{a, b\}$ is true means that either a happens or b happens. Therefore, it represents uncertainty regarding the rating the provider is rated as. In Dempster-Shafer theory, a definition of the Basic Belief Assignment function is first given as follows:

Definition 1. Let Θ be a frame of discernment. A function $m: 2^\Theta \rightarrow [0, 1]$ is a Basic Belief Assignment (BBA) when it satisfies the following two properties [51]:

1. $m\{\emptyset\} = 0$;
2. $\sum_{A \in 2^\Theta} m(A) = 1$.

Then a Belief function is defined as follows:

Definition 2. The Belief (Bel) for a set A ($A \in 2^\Theta$) is defined as the sum of all the assignments of the subsets of A [51]:

$$Bel(A) = \sum_{B \subseteq A} m(B).$$

$Bel(A)$ can be considered as the least probability that the provider is rated as one rating belonging to A . For instance, given a BBA: $m(\{a\}) = 0.9$, $m(\{b\}) = 0$, $m(\{a, b\}) = 0.1$, the belief values can be achieved as: $Bel(\{a\}) = 0.9$, $Bel(\{b\}) = 0$, $Bel(\{a, b\}) = 1$, meaning that the provider will be rated as positive with the probability at least 0.9, as negative with the probability at least 0, and either positive or negative with the probability at least 1.

To combine the evidence from different sources together, a combination rule is defined as follows:

Definition 3. Let Bel_1 and Bel_2 be belief functions over the same frame Θ , with BBAs m_1 and m_2 . Then for each non-empty $A \in 2^\Theta$, the function $m_{12} : 2^\Theta \rightarrow [0, 1]$ that is defined by:

$$m_{12}(\emptyset) = 0,$$

$$m_{12}(A) = \frac{\sum_{i,j:A_i \cap B_j = A} m_1(A_i)m_2(B_j)}{1 - K},$$

where $K = \sum_{i,j:A_i \cap B_j = \emptyset} m_1(A_i)m_2(B_j)$ for all non-empty $A \in 2^\Theta$, is a BBA function. Bel_{12} is the belief function given by m_{12} [51].

As an illustration, suppose two BBAs m_1 and m_2 are defined as follows according to two individual consumer's ratings regarding the same provider:

$$m_1(\{a\}) = 0.9, m_1(\{b\}) = 0.1, m_1(\{a, b\}) = 0,$$

$$m_2(\{a\}) = 0.6, m_2(\{b\}) = 0.2, m_2(\{a, b\}) = 0.2.$$

Then the combined BBA m_{12} can be achieved as follows:

$$m_{12}(\{a\}) = \frac{0.9 \times 0.6 + 0 \times 0.6 + 0.9 \times 0.2}{1 - (0.9 \times 0.2 + 0.1 \times 0.6)} = 0.95,$$

$$m_{12}(\{b\}) = \frac{0.1 \times 0.2 + 0.1 \times 0.2 + 0 \times 0.2}{1 - (0.9 \times 0.2 + 0.1 \times 0.6)} = 0.05,$$

$$m_{12}(\{a, b\}) = \frac{0 \times 0.2}{1 - (0.9 \times 0.2 + 0.1 \times 0.6)} = 0.$$

Correspondingly, we have:

$$Bel_{12}(\{a\}) = 0.95,$$

$$Bel_{12}(\{b\}) = 0.05,$$

$$Bel_{12}(\{a, b\}) = 1,$$

meaning that the provider will be rated as positive with the probability at least 0.95, be rated as negative with the probability at least 0.05, and be rated as either positive or negative with the probability at least 1, according to the evidence (i.e., ratings) from the two sources (consumers). The combining operation satisfies both commutative and

associative laws. Combination of three or more BBAs is achieved by first combining any two BBAs, then combining the product with another BBA, while BBA combination order is not considered here.

Generally, when Dempster-Shafer theory is applied to a real system with K states and N sources ($K \geq 2$, $N \geq 1$), first, a BBA for each source is assigned according to the evidence provided by each distinctive source; second, the combined BBA is computed from the N BBAs by following the combination rule; finally, a belief function that demonstrates the belief of the system states is derived from the combined BBA.

5.3 The Witness Trustworthiness Model

Let's first imagine a face-to-face scenario in reality where the problem of unfair testimonies may also exist. Suppose we are evaluating a person P_1 's reputation but we have not encountered P_1 before. To facilitate our evaluation, we can ask advice from our friend F . Now a new problem arises – how we know F 's advice for P_1 is trustworthy. To solve this problem, we can ask F 's advice for a second person P_2 whom we know whether reputable. If F 's advice for P_2 is similar to our own opinion, then we can consider that F is trustworthy in providing his advice for P_1 . However, only comparing the advice for P_2 may be insufficient. Then we can ask F 's advice for $P_3, P_4, \dots, P_N$, and compare with our own opinions regarding these persons. Finally, we can combine the comparison results together and make a decision on F 's trustworthiness in providing his advice regarding P_1 .

Map the unfair testimony problem in the face-to-face scenario back to that in a reputation system, which includes M providers (denoted as $P_1, P_2, \dots, P_M$). Suppose the provider P_m 's ($1 \leq m \leq M$) reputation is under evaluation by a consumer U who may ask testimonies from the witnesses to evaluate P_m 's reputation more accurately. To determine the trustworthiness of the witness W reporting testimonies, U can use a similar

way. First, U and W 's ratings for the commonly rated providers are compared and the comparison results are combined to achieve W 's personal trustworthiness. In particular, the comparison results are separated into two parts – the comparison results for P_m only and the comparison results for other providers except P_m . The comparison results for P_m are used to calculate W 's personal-local (p-l) trustworthiness value for P_m . The comparison results for other providers except P_m are combined together to calculate W 's personal-global (p-g) trustworthiness value for P_m . The p-l trustworthiness and p-g trustworthiness values are then aggregated together to compute W 's personal trustworthiness value regarding P_m . The separation of p-l and p-g comparison results is to address the scenario where the witness's behavior varies over the providers, such as when collusion attacks exist. If p-l trustworthiness and p-g trustworthiness are treated equally, it is possible that W 's different behaviors for the providers are not differentiated.

Second, to deal with the scenario where U has only a few transactions with the providers or none, W 's ratings are compared with other witnesses' ratings to compute his public trustworthiness regarding P_m . The comparison results for the public trustworthiness are also separated into two parts – public-local (pu-l) and public-global (pu-g). W and other witnesses' ratings regarding P_m are compared to compute W 's pu-l trustworthiness value for P_m . W and other witnesses' ratings regarding other providers are compared and the comparison results are combined to compute W 's pu-g trustworthiness for P_m . Then W 's public trustworthiness for P_m is computed by aggregating W 's pu-l trustworthiness and pu-g trustworthiness together. Finally, we combine W 's personal trustworthiness and public trustworthiness together as his trustworthiness when providing testimonies for P_m . In the parts of computing the personal and public trustworthiness, the rating comparison and combination of comparison results are based on the Dempster-Shafer theory.

5.3.1 Personal Trustworthiness

Suppose in a reputation system, a transaction between U (or W) and a provider P_m happens at time t . After completing the transaction, U (or W) can provide a rating for P_m , which is denoted as $r_{U,P_m}^t = k$ (or $r_{W,P_m}^t = k$)², where k is a value from the integer set $\{1, 2, \dots, K\}$ (K is the number of rating levels the reputation system adopts, and thus $K \geq 2$). We only consider the ratings happened within the most recent time frame Γ (the length of the time frame depends on the reputation systems, for example, it can be one year, half year or one month). In time frame Γ , U has a rating vector R_{U,P_m}^Γ and W has a rating vector R_{W,P_m}^Γ . The ratings in R_{U,P_m}^Γ (or R_{W,P_m}^Γ) are ordered according to the time when they happened³. Then Γ is partitioned into some consecutive elemental windows. The length of an elemental window should adapt to P_m 's behavior change as proposed in [73]. A mapped rating $r_{W,P_m}^{t'}$ is found in R_{W,P_m}^Γ for each rating r_{U,P_m}^t in R_{U,P_m}^Γ . $r_{W,P_m}^{t'}$ should be W 's rating provided at time t' which is closest to time t in the same elemental window. We call $\langle r_{U,P_m}^t, r_{W,P_m}^{t'} \rangle$ as a rating pair. By doing so, U 's ratings are compared to W 's most recently happened ratings to measure the difference between U and W 's ratings. It is also the purpose of we introducing elemental windows, ensuring that the provider behavior changes have small or no impacts on measuring the difference between U and W 's ratings.

The difference d between r_{U,P_m}^t and $r_{W,P_m}^{t'}$ is calculated and expressed as:

$$d = r_{U,P_m}^t - r_{W,P_m}^{t'}. \quad (5.1)$$

It can be seen that d has a total of $2K - 1$ possible values and $-(K - 1) \leq d \leq K - 1$. For each d value, a corresponding α_d is assigned, whose value is initially 0. After the d

²The expression way of a rating in this chapter is different from that in Chapter 3 and 4. In this chapter, a rating is expressed as a numeric number instead of a row vector.

³The expression way of a rating vector in this chapter is different from that in Chapter 3 and 4. In this chapter, a rating vector is an ordered rating sequence according to the time when the ratings happened.

value for a rating pair is calculated, the corresponding α_d value is increased by 1. For example, if $K = 5$, then d will have 9 possible values which are $-4, -3, -2, -1, 0, 1, 2, 3$ and 4 . There will be 9 α_d s which are $\alpha_{-4}, \alpha_{-3}, \alpha_{-2}, \alpha_{-1}, \alpha_0, \alpha_1, \alpha_2, \alpha_3$ and α_4 , respectively. Suppose $r_{U,P_m}^t = 1$ and $r_{W,P_m}^{t'} = 2$ for a rating pair $\langle r_{U,P_i}^t, r_{W,P_m}^{t'} \rangle$. Then d is calculated as -1 . Correspondingly, we increase α_{-1} value by 1.

We count the numbers of all α_d s in all elemental windows. The BBA for W 's trustworthiness regarding P_m from the information source P_m is assigned by using the following way.

First, according to subjective logic [63], for each d , we have:

$$\begin{cases} b_d = \frac{\alpha_d}{\sum_{\sigma=-(K-1)}^{K-1} \alpha_\sigma + 2}, \\ \hat{b}_d = \frac{\sum_{\sigma=-(K-1), \sigma \neq d}^{K-1} \alpha_\sigma}{\sum_{\sigma=-(K-1)}^{K-1} \alpha_\sigma + 2}, \\ u_d = \frac{2}{\sum_{\sigma=-(K-1)}^{K-1} \alpha_\sigma + 2}, \end{cases} \quad (5.2)$$

where b_d is the belief that the proposition that W 's ratings have d level difference from U 's ratings is true, $\hat{b}_d$ is the belief that the proposition that W 's ratings have d level difference from U 's ratings is false, and u_d represents uncertainty. Second, according to the definitions of belief, disbelief, and uncertainty in subjective logic, we can also get:

$$\begin{cases} b_d = m_{W,P_m}(\{d\}), \\ \hat{b}_d = \sum_{y \cap \{d\} = \emptyset} m_{W,P_m}(y), & y \in 2^\Theta \\ u_d = \sum_{y \cap \{d\} \neq \emptyset, y \neq \{d\}} m_{W,P_m}(y), & y \in 2^\Theta \end{cases} \quad (5.3)$$

where $\Theta = \{-(K-1), \dots, 0, \dots, K-1\}$ and m_{W,P_m} is the BBA for W 's trustworthiness according to the information source P_m . Third, solving the equation systems formed by

Equation (5.2) and (5.3), we can get m_{W,P_m} as follows:

$$\begin{cases} m_{W,P_m}(\{d\}) = \frac{\alpha_d}{\sum_{\sigma=-(K-1)}^{K-1} \alpha_\sigma + 2}, \\ m_{W,P_m}(\{-(K-1), \dots, 0, \dots, K-1\}) = \frac{2}{\sum_{\sigma=-(K-1)}^{K-1} \alpha_\sigma + 2}. \end{cases} \quad (5.4)$$

For a provider P_j ($1 \leq j \leq M$) whom U or W does not have any transaction with, the BBA is assigned as:

$$\begin{cases} m_{W,P_j}(\{d\}) = 0, \\ m_{W,P_j}(\{-(K-1), \dots, 0, \dots, K-1\}) = 1. \end{cases} \quad (5.5)$$

After we get BBAs regarding all providers, W 's p-l trustworthiness regarding P_m , T_{W,P_m}^{p-l} , is calculated as the belief of the set $\{d|d_1 \leq d \leq d_2\}$:

$$T_{W,P_m}^{p-l} = Bel_{W,P_m}(\{d|d_1 \leq d \leq d_2\}) = \sum_{d=d_1}^{d_2} m_{W,P_m}(\{d\}), \quad (5.6)$$

where $-(K-1) \leq d_1 \leq d_2 \leq K-1$. We call $d_1 \sim d_2$ as the consumer's *subjective difference tolerance level*. For example, if the consumer considers the witnesses whose testimonies have -1 , 0 , and $+1$ difference from the consumer's personal ratings as trustworthy, he can set $d_1 = -1$, $d_2 = 1$. If the consumer only considers the witnesses who have similar ratings as the consumer's personal ratings as trustworthy, he can set $d_1 = d_2 = 0$. Therefore, the consumer can specify the extent of the subjective difference he can tolerate through setting his own subjective difference tolerance level. It means that WIT does not determine a witness's trustworthiness arbitrarily. It provides the consumer a flexible way to specify his own subjective difference tolerance level and then calculates the witness trustworthiness value accordingly.

To get W 's p-g trustworthiness regarding P_m , the BBAs for the $M - 1$ providers

(P_m is excluded) are combined using the Dempster-Shafer theory combination rule. It is worth pointing out that combining the BBAs from the providers including P_j who has the BBA as Equation (5.5) will have the same effect as combining the BBAs from the providers without P_j . Therefore, combining the BBAs for the $M - 1$ providers is just like combining the BBAs for U and W 's commonly rated providers except P_m . Denote the combined BBA be m_{W,P_m}^{p-g} and the corresponding belief function be Bel_{W,P_m}^{p-g} . W 's p-g trustworthiness regarding P_m , T_{W,P_m}^{p-g} , is calculated as:

$$T_{W,P_m}^{p-g} = Bel_{W,P_m}^{p-g}(\{d | d_1 \leq d \leq d_2\}) = \sum_{d=d_1}^{d_2} m_{W,P_m}^{p-g}(\{d\}). \quad (5.7)$$

To aggregate the p-l trustworthiness and p-g trustworthiness values together, different weights are assigned to them based on $m_{W,P_m}(\{-(K-1), \dots, 0, \dots, K-1\})$. According to Dempster-Shafer theory, $m_{W,P_m}(\{-(K-1), \dots, 0, \dots, K-1\})$ represents the extent of uncertainty (i.e., U does not know whether W is trustworthy or not regarding P_m). Therefore, the larger the $m_{W,P_m}(\{-(K-1), \dots, 0, \dots, K-1\})$ value, the more p-g trustworthiness is necessary to be considered as a complement to the p-l trustworthiness. At the same time, we also want the model to include a learning feature, which is achieved by adopting a sigmod function to map the p-l trustworthiness uncertainty to the complement of the p-l trustworthiness. Then the weights of the p-l trustworthiness ω_{W,P_m}^{p-l} and the p-g trustworthiness ω_{W,P_m}^{p-g} are assigned as follows:

$$\begin{cases} \omega_{W,P_m}^{p-g} = \frac{1}{1 + e^{-\rho \times (m_{W,P_m}(\{-(K-1), \dots, 0, \dots, K-1\}) - 0.5)}}, \\ \omega_{W,P_m}^{p-l} = 1 - \omega_{W,P_m}^{p-g}, \end{cases} \quad (5.8)$$

where ρ decides the step depth of the sigmod function. Figure 5.1 shows the ω_{W,P_m}^{p-g} value changes with $m_{W,P_m}(\{-(K-1), \dots, 0, \dots, K-1\})$ when different ρ values are adopted.

The witness W 's personal trustworthiness regarding P_m , T_{W,P_m}^{per} , is finally computed

Figure 5.1: Global trustworthiness weight changes with local uncertainty

as:

$$T_{W,P_m}^{per} = \omega_{W,P_m}^{p-l} \times T_{W,P_m}^{p-l} + \omega_{W,P_m}^{p-g} \times T_{W,P_m}^{p-g}. \quad (5.9)$$

5.3.2 Public Trustworthiness

W 's public trustworthiness is calculated by comparing W 's ratings with other witnesses' ratings for the providers whom U has no transactions with. Suppose there are other L witnesses, denoted as $W_1, W_2, \dots, W_L$, providing testimonies for a provider P_m whom W provides testimonies for and U has no transactions with. Within the time frame Γ we consider, the witness W_l 's ($1 \leq l \leq L$) rating vector is R_{W_l, P_m}^Γ . Γ is still partitioned into the same elemental windows as in the personal trustworthiness part. In public trustworthiness calculation part, only the last rating provided by each witness regarding P_m in an elemental window is considered⁴. Suppose the last rating value provided by the majority witnesses in an elemental window is r_{maj, P_m}^{last} and the last rating value provided by W is

⁴As it is difficult to find the most closely happened ratings among all the witnesses, we instead use the last rating provided by each witness.

r_{W,P_m}^{last} . The difference d' between r_{maj,P_m}^{last} and r_{W,P_m}^{last} is calculated as:

$$d' = r_{maj,P_m}^{last} - r_{W,P_m}^{last}. \quad (5.10)$$

Similar to d in the personal trustworthiness part, d' also has a total of $2K - 1$ possible values and $2K - 1$ $\alpha'_{d'}$ s are assigned correspondingly. Counting the numbers of all $\alpha'_{d'}$ s in all elemental windows and using the similar way as in the personal-local trustworthiness part, the BBA for W 's pu-l trustworthiness regarding P_m is assigned as:

$$\begin{cases} m'_{W,P_m}(\{d'\}) = \frac{\alpha'_{d'}}{\sum_{\sigma=-(K-1)}^{K-1} \alpha'_{\sigma} + 2}, \\ m'_{W,P_m}(\{-(K-1), \dots, 0, \dots, K-1\}) = \frac{2}{\sum_{\sigma=-(K-1)}^{K-1} \alpha'_{\sigma} + 2}. \end{cases} \quad (5.11)$$

Then W 's pu-l trustworthiness regarding P_m , T_{W,P_m}^{pu-l} , is calculated as:

$$T_{W,P_m}^{pu-l} = Bel'_{W,P_m}(\{d' | d_1 \leq d' \leq d_2\}) = \sum_{d'=d_1}^{d_2} m'_{W,P_m}(\{d'\}). \quad (5.12)$$

The BBA is assigned as Equation (5.5) for a provider whom U has transactions with. Then the BBAs for the $M - 1$ providers (P_m excluded) are combined together. Denote the combined BBA be m_{W,P_m}^{pu-g} and the corresponding belief function be Bel_{W,P_m}^{pu-g} . W 's pu-g trustworthiness regarding P_m , T_{W,P_m}^{pu-g} , is computed as:

$$T_{W,P_m}^{pu-g} = Bel_{W,P_m}^{pu-g}(\{d' | d_1 \leq d' \leq d_2\}) = \sum_{d'=d_1}^{d_2} m_{W,P_m}^{pu-g}(\{d'\}). \quad (5.13)$$

W 's public trustworthiness regarding P_m , T_{W,P_m}^{pub} , is calculated as the weighted sum of W 's pu-l trustworthiness and pu-g trustworthiness:

$$T_{W,P_m}^{pub} = \omega_{W,P_m}^{pu-l} \times T_{W,P_m}^{pu-l} + \omega_{W,P_m}^{pu-g} \times T_{W,P_m}^{pu-g}, \quad (5.14)$$

where the weights for W 's pu-l trustworthiness ω_{W,P_m}^{pu-l} and pu-g trustworthiness ω_{W,P_m}^{pu-g} are assigned using a similar weight assignment function as Equation (5.8):

$$\begin{cases} \omega_{W,P_m}^{pu-g} = \frac{1}{1 + e^{-\rho \times (m_{W,P_m}^{pu-g}(\{-(K-1), \dots, 0, \dots, K-1\}) - 0.5)}}, \\ \omega_{W,P_m}^{pu-l} = 1 - \omega_{W,P_m}^{pu-g}. \end{cases} \quad (5.15)$$

5.3.3 Integration

Finally, W 's trustworthiness regarding P_m is calculated by aggregating the personal trustworthiness and public trustworthiness together. To aggregate them, the personal trustworthiness and public trustworthiness are assigned with different weights based on the uncertainty of personal information, which is achieved as follows. First, the M BBAs obtained in the personal trustworthiness part are combined together by applying the Dempster-Shafer theory combination rule. Denote the combined BBA be m_{W,P_m}^{per} . The weights of the personal trustworthiness ω_{W,P_m}^{per} and public trustworthiness ω_{W,P_m}^{pub} are assigned as follows:

$$\begin{cases} \omega_{W,P_m}^{pub} = \frac{1}{1 + e^{-\rho \times (m_{W,P_m}^{per}(\{-(K-1), \dots, 0, \dots, K-1\}) - 0.5)}}, \\ \omega_{W,P_m}^{per} = 1 - \omega_{W,P_m}^{pub}. \end{cases} \quad (5.16)$$

Finally, the witness W 's trustworthiness regarding P_m , T_{W,P_m} , is calculated as:

$$T_{W,P_m} = \omega_{W,P_m}^{per} \times T_{W,P_m}^{per} + \omega_{W,P_m}^{pub} \times T_{W,P_m}^{pub}. \quad (5.17)$$

5.3.4 A Rating Discounting Scheme

To evaluate provider reputation with more accuracy, we use witness trustworthiness values to discount their reported testimonies. A discounting scheme for binary ratings has been

proposed in [61]. We here extend the scheme for multi-nominal ratings.

Suppose that the time frame Γ includes F time windows, the provider P_m has L witnesses and the witness W_l 's ($1 \leq l \leq L$) aggregated ratings with aging regarding P_m until the last time window are expressed as a row vector A_{W_l, P_m} ⁵. We use b_k , $\hat{b}_k$ and u_k to denote the belief, disbelief and uncertainty that P_m will behave as rating level k in a future transaction. According to subjective logic, we have:

$$\begin{cases} b_k = \frac{A_{W_l, P_m}(k)}{\sum_{k'=1}^K A_{W_l, P_m}(k') + 2}, \\ \hat{b}_k = \frac{\sum_{k'=1, k' \neq k}^K A_{W_l, P_m}(k')}{\sum_{k'=1}^K A_{W_l, P_m}(k') + 2}, \\ u_k = \frac{2}{\sum_{k'=1}^K A_{W_l, P_m}(k') + 2}. \end{cases} \quad (5.18)$$

Summarize the K Equation (5.18), then we can get:

$$\begin{cases} b_k = \frac{A_{W_l, P_m}(k)}{\sum_{k'=1}^K A_{W_l, P_m}(k') + 2}, \\ u_k = \frac{2}{\sum_{k'=1}^K A_{W_l, P_m}(k') + 2}, \end{cases} \quad (5.19)$$

As pointed out in [54], b_k can be directly discounted by the witness's trustworthiness value regarding P_m , then we have:

$$b'_k = T_{W_l, P_m} \times b_k, \quad (5.20)$$

where T_{W_l, P_m} is the trustworthiness value of W_l regarding P_m . Denote the discounted number of ratings for rating level k from W_l regarding P_m be $\hat{A}_{W_l, P_m}(k)$. Substitute $\hat{A}_{W_l, P_m}(k)$ into Equation (5.19), then we have:

$$\begin{cases} b'_k = \frac{\hat{A}_{W_l, P_m}(k)}{\sum_{k'=1}^K \hat{A}_{W_l, P_m}(k') + 2}, \\ u'_k = \frac{2}{\sum_{k'=1}^K \hat{A}_{W_l, P_m}(k') + 2}. \end{cases} \quad (5.21)$$

---

⁵Refer to Equation (3.3) in Chapter 3. For simplicity, we here omit the superscript.

Solve the equation system formed by Equation (5.19), (5.20) and (5.21). $\hat{A}_{W_l, P_m}(k)$ is calculated as:

$$\hat{A}_{W_l, P_m}(k) = \frac{2 \times T_{W_l, P_m} \times A_{W_l, P_m}(k)}{(1 - T_{W_l, P_m}) \times \sum_{k'=1}^K A_{W_l, P_m}(k') + 2}. \quad (5.22)$$

According to Dirichlet Reputation System (DRS) [59] we reviewed in Chapter 4, the probability $Pr(k)$ that P_m will behave as rating level k in the future is calculated as:

$$Pr(k) = \frac{A_{U, P_m}(k) + \sum_{l=1}^L \hat{A}_{W_l, P_m}(k) + 1}{\sum_{k'=1}^K A_{U, P_m}(k') + \sum_{k'=1}^K \sum_{l=1}^L \hat{A}_{W_l, P_m}(k') + K}. \quad (5.23)$$

where A_{U, P_m} is the consumer U 's aggregated rating vector with aging regarding P_m until the last time window. Finally, P_m 's reputation is evaluated as $\hat{E}_{P_m}$:

$$\hat{E}_{P_m} = \sum_{k=1}^K Pr(k) \times k. \quad (5.24)$$

5.4 Experimental Studies

In this section, two sets of experiments are simulated to examine the performance of WIT in coping with the problem of unfair testimonies. In the first set of experiments, a service provision system using a reputation system supporting multi-nominal rating levels is simulated. We first examine the influence of the parameter ρ in the sigmod function on the performance of WIT. Then the witness trustworthiness values are investigated more closely. Finally, we examine the performance of WIT in different scenarios and compare it with Yu and Singh's approach [49] [54] [74]. In the second set of experiments, the performance of WIT for the reputation systems accepting only binary ratings is investigated. We also compare it with the personalized approach [61] in different scenarios. For simplicity, the forgetting factor used in reputation systems is set as 1.

5.4.1 Experimental Studies for Reputation Systems with Multi-Nominal Rating Levels

In this set of experiment, we simulate a service provision system to examine the performance of WIT in addressing the problem of unfair testimonies for reputation systems supporting multi-nominal ratings. We present the experimental setting first, followed by experimental results.

5.4.1.1 Experimental Settings

A service provision system is simulated, including 201 consumers and 100 providers. From the last consumer's point of view, the first 200 consumers are witnesses who will report testimonies regarding the providers to him. The following five types of dishonest witnesses are simulated:

- *D-shifting witnesses* report actual ratings adding D rating levels to the consumer, where D is from the value set $\{i | -(K - 1) \leq i \leq K - 1\}$ (where K is the number of rating levels). But if the actual rating adding D is smaller than 1 or greater than K , the reported rating will be kept as 1 or K ;
- *Not equal witnesses* report to the consumer randomly selected rating levels except the actual ratings;
- *Exaggerated positive (EP) witnesses* [49] report testimonies using the equation: $\hat{r} = \alpha + r - \alpha \times r$, where r is the actual rating, $\hat{r}$ is the reported rating, and α is the exaggerated factor;
- *Exaggerated negative (EN) witnesses* [49] report testimonies using the equation: $\hat{r} = r - \frac{\alpha \times r}{1 - \alpha}$, where r is the actual rating, $\hat{r}$ is the reported rating, and α is the exaggerated factor;

- *Opposite witnesses* report testimonies using the equation $\hat{r} = K - r + 1$.

Five rating levels are simulated. To create a scenario where witness behaviors are diverse and a large number of dishonest witnesses exist, the types of the 200 witnesses are set as shown in Table 5.1.

Witness Index	Witness Type	Witness Index	Witness Type
1-10	-4-shifting	11-20	-3-shifting
21-30	-2-shifting	31-40	-1-shifting
41-50	+4-shifting	51-60	+3-shifting
61-70	+2-shifting	71-80	+1-shifting
81-90	Not equal	91-100	EP, $\alpha = 0.2$
101-110	EP, $\alpha = 0.4$	111-120	EN, $\alpha = 0.2$
121-130	EN, $\alpha = 0.4$	131-140	Opposite
141-200	Honest		

Table 5.1: Simulation witness types

We simulate 10,000 time units and the length of an elemental time window is set as 500 time units (a time unit can be a second, a minute, a day, ..., depending on the reputation systems). We run 100 rounds for each simulation to achieve a statistical accuracy. There are two simulation parameters involved in each round. First, each consumer has a transaction frequency parameter indicating how often he will have a transaction, which is randomly generated in the range of $[0, 1]$. Second, adopting the same setting as in TRAVOS [48], each provider has a major willingness value to simulate his behavior. The major willingness value for each provider is assigned as $0.01 \times \text{provider\_index}$ (i.e., the major willingness value for the 1st provider, 2nd provider, ..., and 100th provider is 0.01, 0.02, ..., and 1, respectively).

More specifically, each round will go through four steps after the round starts. First, it is checked for each consumer whether he will have transaction at current time unit, which is simulated by the consumer's frequency parameter. Second, if the consumer is simulated to have transaction at current time unit, a provider will be randomly selected as

the transaction partner. At the third step, a rating will be generated for this transaction, which is simulated by the involved provider's willingness value. The provider's willingness value is generated from a normal distribution, whose mean is equal to the provider's major willingness value and standard deviation is 0.1⁶. The mapping between willingness values and rating levels is as shown in Table 3.1. Fourth, the witness will report ratings to the consumer⁷. If the witness is an honest one, he will report to the consumer the actual rating generated at the third step. Otherwise, he will report the rating according to his type. After the simulation completes, the mean absolute error (MAE) of the providers' reputation using Equation (4.12) is computed as a measurement of the accuracy of WIT and the comparative approaches.

5.4.1.2 Parameter Value Exploration

Figure 5.2 shows the MAE changes when different ρ values are adopted. The subjective difference level is set as $d_1 = d_2 = 0$. Figure 5.2(a) shows the MAE changes with the number of elemental windows when the dishonest witnesses always report ratings unfairly. Figure 5.2(b) shows the MAE changes with the percentage of unfair testimonies when there are 20 elemental windows.

There are four points noticed according to the results. First, WIT performs much better than Yu and Singh's approach. For example, as shown in Figure 5.2(a), when the two approaches both perform stably after 20 elemental windows, the MAE using Yu and Singh's approach is about 0.8, meaning that there will be an average error of 0.8 rating level for the provider reputation evaluation. But the MAE using WIT is quite close to 0. As shown in Figure 5.2(b), the MAE using WIT is smaller than 0.2 when WIT performs

---

⁶We use 0.1 as standard deviation value to simulate that the provider's behavior does not change much during one elemental window.

⁷In the simulation, we assume that the witnesses report ratings to the consumer. We currently do not discuss how the consumer can collect testimonies in reality.

Figure 5.2: (a) MAE changes with the number of elemental windows; (b) MAE changes with the percentage of unfair testimonies

worst. But the MAE using Yu and Sing's approach is more than 0.8 as a comparison.

Second, no obvious influence on the MAE result can be observed when different ρ values are set, meaning that WIT does not depend on an optimized parameter value setting to achieve a high accuracy. In the following experiments, we choose $\rho = 20$.

Third, the MAE using WIT decreases with the number of elemental windows as shown in Figure 5.2(a). As when more elemental windows are available, more rating pairs can be obtained. The witness trustworthiness can be more accurately evaluated. Therefore, the evaluated provider reputation is closer to the expected reputation and eventually a lower MAE can be achieved.

Fourth, the MAE using WIT presents an arch shape with the increase of the percentage of unfair testimonies as shown in Figure 5.2(b). As with the percentage of unfair testimonies increasing to about 60%, there is more information ambiguity for witness trustworthiness. It is just like in a face-to-face scenario. When a person is honest half of the time, it is difficult to tell whether he will be honest in the future. And with the percentage of unfair testimonies continuously increasing, the ambiguity of the information

becomes smaller. Therefore, MAE decreases with the percentage of unfair testimonies after the percentage is greater than 60%.

5.4.1.3 Witness Trustworthiness Investigation

To take a closer view on the witness trustworthiness, Figure 5.3 shows the witness trustworthiness values for different providers when the consumer sets the subjective difference tolerance level as $d_1 = d_2 = 0$ and $d_1 = -1, d_2 = 1$. In Figure 5.3, x-axis is the index of the 100 providers and y-axis is the witness trustworthiness value achieved through WIT. Generally speaking, three points can be noticed. First, the witness trustworthiness value is specific for different providers as we expect. Second, when different subjective difference tolerance levels are set, the witness trustworthiness value will change accordingly. Third, threshold values can be found to differentiate honest witnesses from dishonest witnesses.

More specifically, we use -1 -shifting witnesses as example to illustrate the first and second noticed points in detail. As Figure 5.3(a) shows, when setting $d_1 = d_2 = 0$, the witness trustworthiness value for -1 -shifting witnesses first decreases with provider index increasing to about 30, and keeps stable for the providers with index 30-80, then decreases with provider index increasing to 100. This result suggests that the witness trustworthiness value is specific for different providers when the witness presents different behaviors over the providers. For the providers with index 1-30, the -1 -shifting witnesses present a more and more dishonest behavior. As a normal distribution with a standard deviation 0.1 is employed to simulate a provider's ratings. The mean value for the normal distribution to generate the provider's ratings is quite small for a provider with a small index (i.e., 1-30th providers). Most of his ratings will be generated as rating level 1 and only a small portion of the ratings will be generated as rating level 2. With the provider index increasing, the portion of the ratings generated as rating level 2 increases. The rating reporting strategy for -1 -shifting witnesses is to report 1 when the actual

Figure 5.3: Witness trustworthiness

rating is 1, otherwise to report 1 level lower. Therefore, with the portion of rating level 2 increasing, the -1 -shifting witnesses will present a more obvious dishonest behavior over these providers.

For the 30-80th providers, the part of the ratings with 1 level difference from most of the provider's ratings is quite stable. For example, most of the 40th provider's ratings will be generated as rating level 2, and a small portion will be generated as rating level 1 or 3. Most of the 60th provider's ratings will be generated as rating level 3, and a small part will be generated as rating level 2 or 4. The percentage of rating level 2 from the 40th provider is similar to the percentage of rating level 3 from the 60th provider. Therefore, the difference between the -1 -shifting witnesses' ratings and consumer's ratings is also similar for the two providers. As a consequence, the witnesses' trustworthiness values for the two providers are close to each other.

For the 80-100th providers, the difference between the -1 -shifting witnesses' ratings and consumer's ratings is more obvious with the portion of the ratings of rating level 5 increasing. Therefore, the witnesses' trustworthiness value decreases with provider index. Similar reasons can explain the change trends of other types of witnesses' trustworthiness values.

When the consumer sets the subjective difference level as $d_1 = -1, d_2 = 1$, the -1 -shifting witnesses can get a high trustworthiness value for all providers as Figure 5.3(a) shows. When the consumer sets $d_1 = -1, d_2 = 1$, it means that the consumer considers trustworthy the witnesses whose ratings have 0 or 1 level difference from the consumer's personal ratings. Therefore, the -1 -shifting witnesses should have a high trustworthiness value. It can be noticed that the -1 -shifting witnesses' trustworthiness values are a bit lower for the 30-80th providers because a quite small portion of the providers' ratings will have 2 level difference from most of their ratings. Therefore, there is a probability that 2 level difference can be observed between the consumer's ratings and the -1 -shifting

witnesses' ratings. Even the consumer sets $d_1 = -1, d_2 = 1$, the trustworthiness for the -1 -shifting witnesses is still a bit lower for the 30-80th providers.

As pointed out in Chapter 3 and 4, we currently use *unfair* testimonies to generally refer to the testimonies that are different from the consumer's personal ratings or the majority witnesses' ratings. The *unfairness* can be a result of intentionally misreporting from malicious witnesses. It can also be a result of subjective difference between the witness and the consumer. But it is difficult to differentiate the unfair testimonies between the two sources. In this chapter, we proposed the concept of *subjective difference tolerance level* to help the consumer to tolerate some extent of subjective difference which is specific for different consumers. For example, some consumers may only consider as trustworthy the witnesses who report similar testimonies as their personal ratings. Some consumers may consider as trustworthy the witnesses whose testimonies have 1-level difference from their personal ratings. By using subjective difference tolerance level, the consumers can specify the subjective difference they can accept in a flexibly way.

The third point noticed from Figure 5.3 is that some threshold values can be found to differentiate honest witnesses from dishonest witnesses. When the consumer sets the subjective difference level as $d_1 = d_2 = 0$, the honest witnesses' trustworthiness value is about 0.45 for the providers with small behavior changes in an elemental window (i.e., the 20-80th providers). Dishonest witnesses' trustworthiness value is smaller than 0.2 for the providers that the witnesses present dishonest behaviors over (e.g., -1 -shifting witnesses' trustworthiness value is about 0.2 for the 25-80th providers, and decreases to 0 for the 81-100th providers). When the consumer sets the subjective difference level as $d_1 = -1, d_2 = 1$, the honest witnesses' trustworthiness value is greater than 0.8 for all providers. -1 -shifting witnesses and $+1$ -shifting witnesses' trustworthiness value is greater than 0.6 for all providers. Other dishonest witnesses' trustworthiness value is smaller than 0.3 for the providers that they present dishonest behaviors over (e.g., -2 -shifting witnesses'

trustworthiness value is about 0.2 for the 41-80th providers and decreases to 0 for the 81-100th providers). According to the simulation results, the threshold values to differentiate honest witnesses from dishonest witnesses can be set as 0.4 when the consumer set the subjective difference tolerance level as $d_1 = d_2 = 0$, and set as 0.6 when $d_1 = -1$, $d_2 = 1$.

5.4.1.4 Robustness Investigation

Figure 5.4 shows the MAE changes in different scenarios. Figure 5.4(a) shows the MAE changes with the percentage of dishonest witnesses. Only some percentage of the first 140 witnesses are randomly selected as dishonest witnesses in this scenario. It can be seen that the MAE using Yu and Singh's approach is more than 0.8 rating level, and the MAE using WIT is significantly lower, which is always smaller than 0.1.

Figure 5.4: (a) MAE changes with the percentage of dishonest witnesses; (b) MAE changes with the number of the providers that the consumer has no transactions with; (c) MAE changes with the number of the providers that the dishonest witnesses are unfair to

Figure 5.4(b) shows the MAE changes with the number of the providers the consumer has no transactions with. The consumer only has transactions with some providers (e.g., 0, 10, ..., and 100 on x-axis means that the consumer has transactions with all providers, 90 providers, ..., and 0 provider, respectively) in this scenario. The MAE using WIT presents an arch shape with a small increase at the tail. The reason behind is similar to that for the MAE result in Figure 5.2(a). The information regarding the witness

trustworthiness is insufficient when the consumer has transactions with some providers only. For example, when the the consumer only has transactions with 50 providers, it is like the scenario where the dishonest witnesses report 50% unfair testimonies and the other 50% are unknown to the consumer. Though the witness's public trustworthiness can be used to complement the 50% unknown information, it is not as accurate as that achieved from using the consumer's personal information directly. As a consequence, the calculated witness trustworthiness is not as accurate as that when more personal information is available. A small increase can be noticed from the MAE corresponding to 90 on x-axis to the MAE corresponding to 100 on x-axis. Since the consumer is totally a newcomer to the reputation system when x-axis is 100. There are no personal ratings available to evaluate the witnesses' trustworthiness. When only public information is used, similar accuracy of witness trustworthiness evaluation cannot be achieved as that when the consumer's personal ratings are available.

Figure 5.4(c) shows the MAE changes with the number of the providers the dishonest witnesses are unfair to. This scenario actually simulates the one where the witnesses' behaviors vary over the providers. It can be seen that the MAE using WIT is always quite small (i.e., < 0.1) no matter how many providers the dishonest witnesses are unfair to.

5.4.2 Experimental Studies for Reputation Systems Accepting Only Binary Ratings

In this set of experiments, we examine the accuracy of WIT for reputation systems accepting binary ratings only and compare it with the personalized approach [61].

5.4.2.1 Experimental Settings

A similar online provision environment as that in the first set of experiments is simulated, including 50 providers and 101 consumers (the first 100 consumers are witnesses from the last consumer's point of view). We simulate one type of dishonest witnesses, who will report the opposite of the actual ratings as testimonies. 10,000 time units are simulated and the length of the elemental window is set as 500 time units. We simulate 100 rounds for each scenario to achieve a statistical accuracy. The simulation follows the same steps as in the first set of experiments. Similar simulation parameters – the frequency of consumers having transaction and the major willingness of providers – are adopted. Compared to the first set of experiments, the difference for this simulation happens at the step of generating the rating for a transaction. In each round of this simulation, a provider is assigned a major willingness value which is randomly generated in the range of $[0, 1]$, indicating the percentage of the positive ratings for this provider (e.g., if the provider's major willingness value is 0.9, 90% of the provider's ratings will be positive). MAE is employed to examine the performance of WIT and the personalized approach. A provider's expected reputation is computed using BRS [39] with the witnesses' actual ratings when all witnesses are supposed to be honest. A provider's evaluated reputation is computed by using BRS with the witnesses' discounted ratings after applying the proposed approach or the personalized approach.

5.4.2.2 Experimental Results

Figure 5.5 shows the MAE changes in different scenarios. Figure 5.5(a) shows the MAE changes with the percentage of unfair testimonies. The unfair testimonies from the witnesses are uniformly distributed among the testimonies for all providers. Figure 5.5(b) shows the MAE changes with the percentage of dishonest witnesses. Figure 5.5(c) shows

the MAE changes with the number of the providers the dishonest witnesses are unfair to, which actually simulates the scenario where the witnesses' behaviors vary over the providers.

Figure 5.5: (a) MAE changes with the percentage of unfair testimonies; (b) MAE changes with the percentage of dishonest witnesses; (c) MAE changes with the number of the providers the dishonest witnesses are unfair to

Two points can be noticed. First, the MAEs achieved from the two approaches are quite close to each other in Figure 5.5(a) and (b). As in the two scenarios, the witnesses behave consistently for all providers. If a witness is dishonest or honest for one provider, he will keep the same behavior for other providers. Since the personalized approach assumes that the witnesses behave consistently for all providers, it can achieve a close MAE result as WIT. Second, as shown in Figure 5.5(c), an obvious difference can be noticed between the MAE using WIT and the MAE using the personalized approach when the witnesses are dishonest for 5 to 45 providers, meaning that when the witnesses behave inconsistently over the providers, WIT performs better than the personalized approach. The scenarios where the witnesses are dishonest for 0 provider and 50 providers actually belong to the scenarios where the witnesses behave consistently. Therefore, the MAEs for the two approaches are close to each other in the two scenarios.

The MAEs for both approaches arise from two aspects – the witnesses' trustworthiness values and the rating discounting scheme. WIT and the personalized approach both

compute the witness trustworthiness values first and further use the witness trustworthiness values to discount their reported ratings. As a similar rating discounting scheme is employed in the two approaches, the MAE difference is caused by witness trustworthiness evaluation. Since the personalized approach calculates a witness's trustworthiness as a general value for all providers, a witness's trustworthiness keeps the same for all providers no matter whether the witness is dishonest for a particular provider. As a consequence, the witness's trustworthiness is less accurately evaluated when the witness behavior varies over the providers. Though WIT can evaluate the witness's trustworthiness specifically for different providers, it still leads to a MAE smaller than 0.2 due to the limitation of the rating discounting scheme.

5.5 Discussions

In this chapter, we proposed a witness trustworthiness model and a rating discounting scheme to address the problem of unfair testimonies. Experimental results show that WIT can effectively address the problem of unfair testimonies with a great extent of flexibility. It is applicable for both the reputation systems supporting multi-nominal ratings and the ones accepting binary ratings only. It is able to work well when the providers and the witnesses present a considerably diverse behaviors. In this section, we will further discuss some potential issues that require more consideration.

First, we introduce “elemental window” to mitigate the influence of provider behavior changes on the accuracy of witness trustworthiness evaluation. But we does not explicitly describe how to divide the elemental windows. In the simulation, we simply set fixed time units as the length of an elemental window and permit a small provider behavior change in each elemental window. But such setting can only be used in the scenarios where the provider behaves consistently along the time or changes his behavior with a

small variation. In reality, a better way to set the length of the elemental window is to adapt to the frequency of the provider behavior change. It can be achieved by using a provider behavior change detector from time domain [73].

Second, though WIT does not depend on the assumption that the witness behaves consistently for all providers, it assumes that the witness behaves consistently along time. If the witness is honest during some time windows and dishonest during some other time windows, WIT will treat the witness's behavior from a global point of view. For example, suppose there are 100 time windows considered and a witness is honest during the first half 50 time windows and dishonest during the last 50 time windows. WIT will consider the witness as half honest during the whole time frame rather than compute exactly in what time window the witness is honest. An option to resolve this problem is to leave it to the reputation system by discounting the witness's reported ratings with both the witness's trustworthiness value and a time decay factor as proposed in [39] [59].

The third issue requiring more consideration is related to the problem of collusion attacks. Though WIT does not assume that the witnesses behave consistently for all providers, it may not work when some special collusion attacks exist. For example, if a large amount of witnesses collude together to report ratings unfairly for one particular provider only and the consumer himself coincidentally has no transactions with this provider, then using only public information may produce a wrong witness trustworthiness result. The pu-l trustworthiness value for the actually dishonest witnesses will be high as the majority witnesses are dishonest. The pu-g trustworthiness value for the actually dishonest witnesses will also be high as the colluded witnesses are honest for other providers. As a consequence, the colluded dishonest witnesses will get a high trustworthiness value for the provider that the colluded dishonest witnesses are against, and the actual honest witnesses will wrongly get a low trustworthiness value.

5.6 Summary

In this chapter, we proposed WIT – a witness trustworthiness model, and a discounting scheme to discount the witness’s reported ratings according to the calculated witness trustworthiness value. We utilize Dempster-Shafer theory that has not been used to model witness trustworthiness previously, incorporate local and global, personal and public information to model witness trustworthiness, and propose a flexible way to consumers to specify their subjective difference tolerance levels. Compared with the proposed clustering based filtering approaches in Chapter 3 and 4, and the approaches reviewed in Chapter 2, WIT has the following advantages:

- It does not depend heavily on an extremely accurate parameter setting to address the problem of unfair testimonies. As the experimental results show, the value of the only parameter ρ almost has no effect on the performance of WIT.
- It does not assume that the witnesses behave consistently over the providers to achieve a good performance. It can model the witness trustworthiness particularly for different providers as experimental results present. Therefore, it can resist collusion attacks to a good extent.
- It allows the consumer to specify his own subjective difference tolerance level instead of making an arbitrary decision on which witnesses are trustworthy. In most approaches we reviewed in Chapter 2, the two-stage clustering filtering approach proposed in Chapter 3, and the iCULB approach proposed in Chapter 4, we do not differentiate the unfair testimonies caused by subjective difference from malicious reporting. To allow consumers to accept some “unfair” testimonies caused by subjective difference, WIT provides subjective difference tolerance level to the consumers to specify the extent of subjective difference they can accept. By using

subjective difference tolerance level, some “unfair” testimonies caused by subjective difference can be considered as acceptable and differentiated from the unfair testimonies caused by malicious reporting.

Chapter 6

Conclusions and Future Work

Trust management has been a concern in MAS for a long time. With the popularity of online service provision systems, trust management has received a lot of attention to secure online service provision environment. Reputation systems have been proposed as a mechanism for trust evaluation, and have been an important part for the success of online service provision systems. But their reliability has been seriously jeopardized due to the existence of unfair testimonies. Mitigating the adverse effect of unfair testimonies is an urgent issue to be addressed to build a reliable reputation system. This chapter summarizes our effort in dealing with the problem of unfair testimonies and is structured as follows. Section 6.1 reviews the contributions of the thesis work. Section 6.2 presents the potential applications the work can be applied to. Section 6.3 discusses the possible directions the thesis work can be continued in the future. Section 6.4 concludes this thesis.

6.1 Contributions

This thesis devotes itself to building reliable reputation systems in the presence of unfair testimonies. To accomplish this goal, this thesis has made the following contributions.

- As the first attempt to address the problem of unfair testimonies, we proposed a two-stage clustering filtering approach in Chapter 3. According to the categorization proposed in Chapter 2, it can be classified as an approach adopting filtering mechanism by using the local, personal and public information. The two-stage clustering filtering approach imagines the ratings from the witnesses and the consumer as rating vectors in a K -dimensional space, where K is the number of rating levels the reputation systems adopt. Then a two-stage clustering approach based on the hierarchical clustering method is applied to cluster the rating vectors. In the first stage, the rating vectors with the closest similarity are clustered together. In the second stage, the rating vectors whose difference does not exceed the threshold are continuously clustered together. Then the ratings which are not in the same cluster as the consumer's personal ratings or the majority witnesses' ratings are identified as unfair testimonies and discarded. The remaining ratings will be passed to reputation systems for provider reputation evaluation.

Compared with existing work, the two-stage clustering filtering approach can support reputation systems with multi-nominal rating levels, which is a major limitation of the existing approaches. As it considers local, personal and public information, and assign personal information a higher weight, it can work when majority testimonies are unfair when personal information is available. Experimental results show that the two-stage filtering approach can be integrated into a reputation system easily and outperform other existing approaches such as the iterated filtering approach and TRAVOS.

- In Chapter 4, we proposed an integrated clustering filtering approach (iCLUB) to improve the two-stage clustering filtering approach proposed in Chapter 3. Compared with the two-stage clustering filtering approach, iCLUB further incorporates global information to filter unfair ratings. To address the scenario where the two-

stage clustering approach cannot work, for example, majority ratings are unfair and the consumer has no personal ratings regarding the target provider, iCLUB uses the global information to find a set of common honest witnesses who are honest for the providers the consumer has encountered before. And only the witnesses who are in the same cluster as the consumer or in the cluster with the greatest number of common honest witnesses are considered as honest for the target provider.

Experimental results show that iCLUB can address the problem of unfair testimonies in various scenarios especially when collusion attacks exist. It can be easily integrated with a reputation system, for example, DRS and BRS. It outperforms the related work, such as the iterated filtering approach and TRAVOS, in both the accuracy of filtering unfair testimonies and evaluating provider reputation.

- Adopting modeling mechanism, we proposed a witness trustworthiness model (WIT) based on Dempster-Shafer theory in Chapter 5. WIT considers personal-local, personal-global, public-local, and public-global information to model a witness's trustworthiness which is specific for different providers. The calculated witness trustworthiness value is further used to discount the witness's testimonies to evaluate the provider's reputation more accurately. Compared with the existing approaches, WIT does not assume that the witnesses behave consistently over all providers. It can output a witness trustworthiness value which is specific for different providers. Therefore, it can counter collusion attacks to a good extent. Experimental results show that WIT can model the witness trustworthiness effectively and flexibly. It can evaluate provider reputation more accurately and outperform the comparative approaches, for example, Yu and Sing's approach and the personalized approach, with a significantly lower mean absolute error regarding provider reputation evaluation. WIT also has the remarkable feature that it provides consumers a flexible way to specify their own subjective difference tolerance levels. As we have mentioned in

Chapter 3, unfair testimonies can be the result of malicious witnesses misreporting ratings intentionally, and they also can be the result of the subjective difference between the consumer and witnesses. Most of the existing work does not differentiate the two sources of unfair testimonies and use an arbitrary threshold to identify the unfair ratings no matter whether they are caused by subjective difference. However, in reality, some consumers may consider the ratings with some extent of subjective difference as acceptable. To deal with such scenarios, the concept of subjective difference tolerance level is proposed in WIT to allow consumers to accept some “unfair” testimonies caused by subjective difference by specifying the extent of subjective difference they can accept.

Another significant merit of the three approaches proposed in this thesis is that they are decoupled from the reputation systems. They are not explicitly designed for a particular reputation system. Although DRS and BRS are employed as examples to show how the approaches can be integrated with a well-developed reputation system, other reputation systems can be supported as long as the systems can provide the ratings as the approaches require. This makes them a more generic solution to address the problem of unfair testimonies for reputation systems. As a summary, we list the mechanisms the three approaches adopt, the information they consider, and the capabilities they have in Table 6.1, 6.2 and 6.3 by comparing with the approaches we reviewed in Chapter 2, respectively.

6.2 Potential Applications

Though the thesis work focuses on addressing the problem of unfair testimonies in online service provision systems, it is reiterated that the proposed research can be applied to many application domains of MAS which are not limited to online service provision

Approach	Filtering	Modeling	Learning	Incentive
Iterated Filtering	✓			
Entropy-Based Filtering	✓			
Cluster Filtering	✓			
Multi-Layer Cognitive Filtering	✓	≈		
TRAVOS	≈	✓		
Yu and Singh	≈	✓		
Bayesian Learning			✓	
CHI-Score	≈			✓
Personalized	≈	✓		
Trust-Based Incentive	≈	≈		✓
Two-Stage Clustering Filtering	✓			
iCLUB	✓			
WIT	≈	✓		≈

Table 6.1: The mechanism our proposed approaches adopt

Approach	Personal	Public	Local	Global
Iterated Filtering		✓	✓	
Entropy-Based Filtering	✓	✓	✓	
Cluster Filtering		✓	✓	
Multi-Layer cognitive filtering	✓			✓
TRAVOS	✓			✓
Yu and Singh	✓		✓	
Bayesian Learning	✓			✓
CHI-Score		✓	✓	
Personalized	✓	✓		✓
Trust-Based Incentive	✓	✓		✓
Two-Stage Clustering Filtering	✓	✓	✓	
iCLUB	✓	✓	✓	✓
WIT	✓	✓	✓	✓

Table 6.2: The information our proposed approaches consider

systems only. Some typical domains include e-commerce systems and cognitive radio networks.

6.2.1 E-Commerce Systems

As pointed out in the motivation part, our main goal of the thesis work is to address the problem of unfair testimonies for the reputation system in online service provision

Approach	Majority	Flooding	Lack	Provider Varying	Witness Varying
Iterated Filtering			✓	✓	✓
Entropy-Based Filtering	✓	✓		✓	✓
Clustering Filtering		✓	✓		✓
Two-Layer Cognitive Filtering	✓	✓			
TRAVOS	✓	✓			
Yu and Singh	✓	✓			✓
Bayesian Learning	✓	✓			
CHI-Score		✓	✓		✓
Personalized	✓	✓	✓	✓	
Trust-Based Incentive	✓	✓	✓	✓	
Two-Stage Clustering Filtering	✓	✓	✓	✓	✓
iCLUB	✓	✓	✓	✓	✓
WIT	✓	✓	✓	✓	✓

Table 6.3: The capabilities our proposed approaches have

systems. With respect to this, e-commerce system is a direct application domain. With the development of Internet and other network techniques, people have changes the conventional way to conduct business. As individuals and organizations bring their business through internet, buyers can make orders online from anywhere in the world. Benefits and opportunities are brought with these changes [75] [76], and e-commerce becomes increasingly popular as a consequence. However, challenges also follow the benefits. The problem of unfair testimonies is a notable one, which will influence the accuracy of seller reputation evaluation and therefore harm buyers' benefits.

Our work in the thesis can be used to cope with the problem of unfair testimonies in e-commerce as one of our publications¹ shows. When applying our work into an e-commerce system (for example, eBay), the ratings regarding the sellers are shared through a central server which can perform the task of filtering unfair testimonies or modeling witness trustworthiness. Then more accurate results of seller reputation evaluation can be made available to buyers to facilitate them to select transaction partners.

¹Conference paper [2] in author's publication list.

6.2.2 Cognitive Radio Networks

In the traditional fixed spectrum assignment scheme, there exist the problems of spectrum scarcity and inefficient spectrum utilization [77]. Cognitive radio network (CRN) is proposed to address these problems [78]. More specifically, there are two types of users in a CRN – primary users (PU) and secondary users (SU). When PUs occupy the spectrum, the SUs should not interfere the PUs. And when the spectrum is free for using, the SUs can transmit their own packet. Therefore, CRN working efficiency is based on the spectrum sensing and spectrum state decision. The SUs should have the capability of sensing the actions of PUs and adapt to their actions, which can be more efficiently achieved by multiple SUs sharing sensing information and cooperation together [79]. However, new challenges follow the cooperation. The problem of unfair testimonies is one of them. Some malicious SUs may misreport the sensing information regarding the PUs to mislead other SUs' decisions.

Our work in the thesis can be used to cope with the problem of unfair testimonies in CRN as one of our publications² shows. When applying our work into a CRN environment, each SU will report the testimonies regarding the accuracy of other SUs sensing the spectrum usage when being asked. Then the SU asking testimonies can analyze the collected testimonies using the approaches proposed in the thesis, aggregate the fair testimonies, and make a decision on the availability of the spectrum.

6.3 Future Work

For the future, we plan to expand our current research in both improving witness trustworthiness model and application area extension.

---

²Conference paper [5] in author's publication list.

6.3.1 Improving the Witness Trustworthiness Model

The witness trustworthiness model can be extended from the aspects of elemental window size setting, building incentive mechanism to promote honesty, collecting testimonies and dealing with subjective difference.

6.3.1.1 Elemental Window Size Setting

In Chapter 5, we proposed a witness trustworthiness model (WIT). Elemental windows are used to mitigate the influence of provider behavior changes. But we did not explicitly point out how to set the size of an elemental window. In the experimental studies, we used fixed length as the size of an elemental window. But in reality, more factors need to be considered.

For instance, the determination of the size of a time window proposed in [45] is based on the frequency of the fair ratings reporting for a particular provider. To make the time window contains average k fair ratings, a smaller time window will be adopted when the system carries more ratings to the given provider. Though the purpose of setting time windows in [45] is different from that of setting elemental windows in WIT, it can be seen that some factors regarding the level of system activity are able to be employed to the determination of elemental window size [19].

The extent and frequency of provider behavior changes are also factors determining the proper size of an elemental window [19]. In the experimental studies we conducted in Chapter 5, we simulated a small provider behavior change during an elemental window by using a low standard deviation value to show that WIT can still work accurately when provider behaviors vary slightly. However, as we have discussed in Chapter 5, the setting of elemental window size also needs to adapt to the provider behavior changes [73].

For future work, we will explore an effective way to determine an appropriate size of

an elemental window by considering both the the level of activity in the system and the extent and frequency of provider behavior changes. Adjusting thes factors based on the collected testimonies, the most appropriate elemental window size will be selected and adapted to achieve best performance of witness trustworthiness evaluation [19].

6.3.1.2 Building Incentive Mechanisms to Promote Honesty

According to the approach categorization we introduced in Chapter 2, our developed witness trustworthiness model belongs to the modeling mechanism category among the approaches addressing the problem of unfair testimonies. But it is actually compatible to the category of incentive mechanism after extension.

Generally speaking, the aim of an incentive mechanism is to achieve an equilibrium (a stationary point in the system) where the best outcomes are obtained for the witnesses' actions of providing truthful testimonies regarding providers. According to the categorization proposed in [64], the incentive mechanism approaches can be divided into three categories, which are side payment mechanism approaches [57] [64] [66] [80] [81] [82], credibility mechanism approaches [83] [84], and trust revelation mechanism approaches [85] [86].

In the side payment mechanism approaches, side payments are offered to consumers that provide fair testimonies regarding their transaction partners. For example, in [80], trustful ratings from consumers are elicited by offering rebates of a consumer's periodic membership fee. A Nash equilibrium can be achieved when trustful ratings are provided by consumers. However, the side payment mechanism approaches cannot work well when majority consumers provide unfair ratings as each dishonest consumer will receive a reward in such case. In other words, the honest consumers that do not provide similar ratings as other majority consumers will not be rewarded and therefore discouraged to report trustfully in future [19]. The side payment mechanism approaches also have the problem that they may require additional overhead to control the side payments. For example,

in [81] [82], a central entity is required to make sure that the mechanism pays off and receives the same monetary amount.

Compared with the side payment mechanism approaches, credibility mechanism approaches elicit consumers or providers to report trustful ratings by measuring their credibility instead of giving instant payment. Credibility mechanism approaches are designed based on the assumption that consumers or providers are more likely to transact with the credible ones. Therefore, a consumer or a provider proved to be more credible is more likely to attract more transaction partners. In credibility mechanism approaches, the transaction partners (i.e., a consumer and a provider) will report ratings regarding the behavior of each other. If the two reported ratings do not match each other, then the credibility of the two partners will be decreased. As a consequence, consumers will report trustful ratings for the purpose of keeping up their credibility. However, it also brings the problem to the front that an honest consumer or provider will lose their credibility if they unfortunately meet with a dishonest partner. The honest consumer or provider cannot gain credibility even if he provides trustful ratings. The credibility mechanism approaches also have the problem that they cannot address the scenario where a provider and a consumer collude together to increase the credibility of each other [64].

Different from side payment mechanism approaches and credibility mechanism approaches, trust revelation mechanism approaches are designed to elicit reporting trustworthiness value trustfully regarding a consumer or a provider put by another. For example, in [86], a trust revelation mechanism is proposed for the scenario of task allocation. In this scenario, a consumer needs to make decisions on which providers he should allocate his task to. First, the consumer reports to the center a set of tasks to be allocated. Then each provider reports his cost and valuation for completing these tasks, as well as the trustworthiness value it places on other providers. The center calculates the trustworthiness of each provider based on the reported trustworthiness value of the provider

from other providers. Then the center determines the optimal allocation and payment according to Vickry-Clarke-Groves (VCG) mechanism. The trust revelation mechanism approaches suffer from the problem of collusion attacks. The providers may provide inaccurate trustworthiness information regarding other providers to increase or decrease the probability that another provider receives a good allocation [64].

To address the problems of existing incentive mechanism approaches mentioned above, we are trying to develop an incentive mechanism approach that has the following features as suggested in [19]. First, the witness who has a higher trustworthiness value should be rewarded with a higher side payment. Second, not all witnesses will receive rewards. A witness with a quite low trustworthiness value should be punished. Third, the side payment for the honest witnesses is from the punished witnesses to achieve a Nash Equilibrium. Though the trust-based incentive mechanism approach [64] we reviewed in Chapter 2 has the above features, it may not reward consumers correctly as it inherits the disadvantages of the personalized approach. Therefore, we are trying to design an incentive mechanism approach that has the above features and avoids the disadvantages of trust-based incentive mechanism approaches by incorporating the proposed witness trustworthiness model in the future.

6.3.1.3 Collecting Testimonies

No matter in the clustering based filtering approach proposed in Chapter 3 and 4, or the witness trustworthiness model proposed in Chapter 5, we have set aside the problem of testimony collection with the assumption that the consumer can somehow collect the ratings. But the problem of testimony collection needs more exploration in reality.

In a centralized reputation system, the ratings are shared among all consumers. The consumers can access the shared ratings directly. The center will take responsibilities of addressing the problem of unfair testimonies and evaluating provider reputation using

the shared ratings. But some issues still need a careful consideration. For example, a witness who only provides a few ratings may not be considered as an eligible witness as the ratings provided by him may not reflect the witness's real characteristics (i.e., trustworthy or untrustworthy) or the provider's real behavior [87].

In a decentralized reputation system, one of the challenging problems in collecting testimonies is how to balance the benefits from the greater amount of information when the pool of witnesses is large and the concerns regarding the greater change of unreliable information sources when the pool is not small [19]. There are some existing work discussing how to collect testimonies in a decentralized reputation system [48] [49]. A common way is to build a social network for each consumer to collect testimonies. The consumer can request ratings from the witnesses in his network. But there are some issues requiring careful consideration. For example, the depth (i.e., how deep a consumer can search for testimonies) and width (i.e., how many witnesses can be directly connected to a consumer) of the social network need to be set carefully. If the size of the witnesses connected to a consumer is too large, the computational cost for maintaining and updating of witness trustworthiness will be increased. A smaller size may decrease the computation cost, but it may happen that the witnesses have insufficient experience with providers [19] [87]. Therefore, a trade-off between the size of the computation cost and information sufficiency needs to be explored. How to discount and aggregate the ratings collected from the indirectly connected witnesses is also an issue requiring consideration.

6.3.1.4 Dealing with Subjective Difference

For the clustering based filtering approaches proposed in Chapter 3 and 4, we did not differentiate the unfair testimonies caused by intentionally malicious reporting from the unfair testimonies caused by subjective difference. For the witness trustworthiness model proposed in Chapter 5, we introduced the concept of subjective difference tolerance level to

allow the consumers to accept some extent of “unfair” testimonies caused by subjective difference. But the problem of handling subjective difference needs more exploration. If two consumers have subjective difference with respect to the service accepted, one consumer’s ratings may seriously mislead another consumer’s decision [88]. For example, in an e-commerce environment, a seller may be rated with a low rating from a buyer if the seller delivers the item late. But if a second buyer does not consider delivery speed as significantly as the first buyer, then the first buyer’s low rating will mislead the second buyer’s evaluation on the seller’s reputation [19].

To handle the problem of subjective difference better, the concept of subjective difference tolerance level can be extended. For example, in the Bayesian learning approach [2], a consumer can learn the witnesses’ behaviors on reporting ratings and infer the ratings back to the providers’ real behaviors. In the two-layer cognitive filtering approach, a consumer can categorize the witnesses into credible, optimistic, pessimistic and debitable witnesses. Though the two approaches have the disadvantage that they are designed based on the assumption that the witnesses behave consistently over all providers, the problem of subjective difference can be addressed better if we can integrate the advantages of WIT with the merits of learning witness behaviors and inferring ratings back to real provider behaviors.

6.3.2 Applying to Other Applications

Beyond the applications we discussed in 6.2, the thesis work can contribute to other applications after extension and exploration in the future.

6.3.2.1 Review Systems

Review systems are such systems where consumers can post the numerical ratings with textual comments, images or even videos to rate the quality of the service they received.

For example, the popular website, yelp (<http://www.yelp.com/>), is a review system where users can post feedbacks regarding the restaurants they have gone for and the hotels they have resided in. The notable movie review website, IMDB (<http://www.imdb.com/>), is also a review system where users can rate the movies they have watched.

The review systems have a significant difference from the online service provision systems that the proposed approaches in the thesis are designed for. In our thesis work, we have an assumption that the consumers have multiple transactions with a single provider. This assumption can be met in some systems, such as an e-commerce system, a web-service provision system, or a P2P file sharing system. But usually a review system cannot meet this assumption. For example, in IMDB, the possibility that a user posts multiple reviews for a single movie is very low. Therefore, to apply our approaches to review systems, it needs more consideration. After careful modification and extension, the proposed witnesses trustworthiness model will be able to be applied to review systems. For example, as one of our publications³ shows, integrating the witnesses trustworthiness model and an unfair rating detector in time domain can mitigate the influence of unfair testimonies for an online product review system. We are also working on designing other approaches to address the problem of unfair testimonies in review systems. As one of our publications⁴ shows, fuzzy logic is employed to mitigate the influence of unfair testimonies for an online product review system.

6.3.2.2 Ad-Hoc Networks

After extension, the thesis work can be applied to vehicular ad-hoc networks (VANET) [89] where the latest information regarding road conditions can be communicated among the agents in vehicles to improve road safety. Due to their self-interested nature, the agents may misreport road condition related information to mislead other agents' decision. To

---

³Journal paper [2] in the publication list.

⁴Conference paper [1] in the publication list.

apply the thesis work in VANET, one major necessary extension is related to the problem of testimony collection. In VANET, the environment may change continuously and rapidly and an agent can report information regarding various events. A reliable reputation model should be able to capture the information diversity and environment dynamism rapidly. Therefore, the determination that from whom the agent should collect testimonies is significantly important in order to make a quick decision for the time-sensitive tasks in the dynamic VANET environment [19].

After extension, the thesis work can also be applied to collaborative intrusion detection network (IDN) [90], where an intrusion decision can be made before the worm infects most of the network hosts by propagating alerts among participants. To propagate alerts efficiently and quickly, the collective knowledge and experience are shared amongst the members of the intrusion detection system. Considering each host sharing information as a consumer, our work can be applied to IDN by modeling the trustworthiness of host sharing information. Similarly, the problem of collecting testimonies needs careful consideration as a decentralized reputation system is usually adopted in IDN.

6.4 Conclusions

In this thesis, we first developed a two-stage clustering filtering approach to tackle the problem of unfair testimonies for reputation systems. Then we improved the two-stage clustering filtering approach to an integrated clustering filtering approach by considering more information. Using a different mechanism, we further proposed a witness trustworthiness model based on Dempster-Shafer theory to model witness trustworthiness effectively and flexibly by adapting to the consumer specified subjective difference tolerance level. Experimental results show that the thesis work significantly mitigates the adverse effect of unfair testimonies and contributes to building reliable reputation systems.

For future work, we plan to extend our work to make it more applicable to real-world environments. If we are able to adapt our work to make it applicable in a realistic system, such as eBay and CRN, and extend our work to review systems and some ad-hoc networks, our work will contribute to eliminating users' hesitance of using online service provision systems with the merits of secure, convenience, and flexibility.

Author's Publications

Journal Papers

1. **Siyuan Liu**, Jie Zhang, Chunyan Miao, Yin-Leng Theng, and Alex C. Kot, “An Integrated Clustering-Based Approach to Filtering Unfair Multi-Nominal Testimonies”, *Computational Intelligence*, in press.
2. Yuhong Liu, Yan Sun, **Siyuan Liu**, and Alex C. Kot, “Securing Online Reputation Systems through Trust Modeling and Temporal Analysis”, *IEEE Transactions on Information Forensics and Security*, in press.

Conference Papers

1. **Siyuan Liu**, Han Yu, Chunyan Miao, and Alex C. Kot, “A Fuzzy Logic Based Reputation Model against Unfair Ratings”, in *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems, (AAMAS'2013)*, Saint Paul, USA, 2013.
2. **Siyuan Liu**, Alex C. Kot, Chunyan Miao, and Yin-Leng Theng, “A Dempster-Shafer Theory Based Witness Trustworthiness Model to Cope with Unfair Ratings in E-Marketplace”, in *Proceedings of the 14th Annual International Conference on Electronic Commerce (ICEC'2012)*, Singapore, 2012 (**Best Paper Award**).
3. **Siyuan Liu**, Alex C. Kot, Chunyan Miao, and Yin-Leng Theng, “A Dempster-Shafer Theory Based Witnesses Trustworthiness Model”, in *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems, (AAMAS'2012)*, Valencia, Spain, 2012.
4. **Siyuan Liu**, Jie Zhang, Chunyan Miao, Yin-Leng Theng, and Alex C. Kot, “iCLUB: An Integrated Clustering-Based Approach to Improve the Robustness of Reputation Systems”, in *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'2011)*, Taipei, Taiwan, 2011.

5. Han Yu, **Siyuan Liu**, Alex C. Kot, Chunyan Miao, and Cyril Leung, “Dynamic Witness Selection for Trustworthy Distributed Cooperative Sensing in Cognitive Radio Networks”, in *Proceedings of the 13th International Conference on Communication Technology (ICCT'2011)*, JiNan, China, 2011 (**Best Paper Award**).
6. **Siyuan Liu**, Chunyan Miao, Yin-Leng Theng, and Alex C. Kot,, “A Clustering Approach to Filtering Unfair Testimonies for Reputation Systems”, in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'2010)*, Toronto, Canada, 2010.

Submitted Papers

1. **Siyuan Liu**, Alex C. Kot, Yin-Leng Theng, Chunyan Miao, and Yan Sun, “A Dempster-Shafer Theory Based Witness’s Trustworthiness Model for Reputation Systems”, *submitted to IEEE Transactions on Dependable and Secure Computing*, under revision.

Bibliography

- [1] A. Whitby, A. Jøsang, and J. Indulska, “Filtering out unfair ratings in bayesian reputation systems,” *ICFAIN J. Management Res.*, vol. 4, no. 2, pp. 48–64, 2005.
- [2] K. Regan, P. Poupart, and R. Cohen, “Bayesian reputation modeling in e-marketplaces sensitive to subjectivity, deception and change,” in *Proceedings of the 21st National Conference on Artificial Intelligence*, pp. 206–212, 2006.
- [3] G. Weiss, *Peer-to-peer: harnessing the power of disruptive technologies*. O’Reilly & Associates Inc., 2001.
- [4] I. Foster and C. Kesselman, *The grid, blueprint for a new computing infrastructure*. Morgan Kaufmann Inc., 1998.
- [5] A. Oram, *Multiagent systems: a modern approach to distributed artificial intelligence*. The MIT Press, 2000.
- [6] T. Berners-Lee, J. Hendler, and O. Lassila, “The semantic web,” *Scientific American*, vol. 285, pp. 34–43, 2001.
- [7] S. A. Mcilraith, T. C. Son, and H. Zeng, “Semantic web services,” *IEEE Intelligent Systems*, vol. 16, pp. 46–53, 2001.
- [8] A. Rogers, E. David, J. Schiff, and N. R. Jennings, “The effects of proxy bidding and minnum bid increments within eBay auctions,” *ACM Transactions on the Web*, vol. 1, no. 2, 2002.
- [9] N. Sadeh, *M-commerce: technologies, sevicees, and business models*. Wiley Computer Publishing, 2002.
- [10] H. Schmeck, T. Ungerer, and L. C. Wolf, “Trends in network and pervasive computing,” *Lecture Notes in Computer Science*, vol. 2299, 2002.
- [11] J. O. Kephart and D. M. Chess, “The vision of autonomic computing,” *IEEE Computer*, vol. 36, pp. 41–50, 2003.

- [12] J. H. Smith, *Trusting the avator*. <http://jonassmith.dk/weblog/uploads/TrustingtheAvatardraftsubmitted.pdf>, 2004.
- [13] P. Beaumont and B. Chaib-draa, "Multiagent coordination techniques for complex environments: the case of a fleet of combat ships," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 37, no. 3, pp. 373–385, 2007.
- [14] T. Máhr, J. Srour, M. de Weerdt, and R. Zuidwijk, "Can agents measure up? A comparative study of an agent-based and on-line optimization approach for a drayage problem with uncertainty," *Transportation Research Part C: Emerging Technologies*, vol. 8, no. 1, pp. 99–119, 2010.
- [15] S. Basak, N. Modanwal, and B. D. Mazumdar, "Multi-agent based disaster management system: a review," *International Journal on Computer Science and Technology (IJCST)*, vol. 2, no. 2, pp. 343–348, 2011.
- [16] C. Kadushin, *Understanding social networks : concepts, theories, and findings*. Oxford University Press, 2012.
- [17] M. Wooldridge and N. R. Jennings, "Intelligent agents: theory and practice," *The Knowledge Engineering Review*, vol. 10, no. 2, pp. 115–152, 1995.
- [18] K. S. Barber and J. Kim, "Soft security: isolating unreliable agents from society," *Lecture Notes in Artificial Intelligence*, pp. 224–233, 2003.
- [19] J. Zhang, *Promoting honesty in electronic marketplaces: combining trust modeling and incentive mechanism design*. Ph.D thesis, University of Waterloo, 2006.
- [20] J. Weng, *Building robust trust model for multi-agent systems*. Ph.D thesis, Nayang Technological University, 2008.
- [21] D. Gambetta, *Trust: making and breaking cooperative relations*. Basil Blackwell, 1988.
- [22] Y. Liu, Y. Sun, S. Liu, and A. C. Kot, "Securing online reputation systems through trust modeling and temporal analysis," *IEEE Transactions on Information Forensics and Security*, in press.
- [23] Y. blog, *Holy Nyans! 60 hours per minute and 4 billion views a day on YouTube*. <http://youtube-global.blogspot.com/2012/01/holy-nyans-60-hours-per-minute-and-4.html>, 2012.
- [24] J. Moire, *Facebooks 845 million members may not all visit the site*. <http://www.allfacebook.com/facebook-users-2012-02>, 2012.

- [25] A. Noce and C. Peters, "Barriers to electronic commerce in Canada: a size of firm and industry analysis," in *Proceedings of the Economic Conference 2005 - Growth, Wealth and Sustainability*, pp. 1–25, 2005.
- [26] E. M. Maximilien and M. P. Singh, "Agent-based trust model involving multiple qualities," in *Proceedings of the 4th International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 519–526, 2005.
- [27] J. Sabater and C. Sierra, "Social regret, a reputation model based on social relations," *SIGecom Exchanges*, vol. 3, no. 1, pp. 44–56, 2002.
- [28] T. Huynh, N. R. Jennings, and N. Shadbolt, "An integrated trust and reputation model for open multi-agent systems," in *Proceedings of the 16th European Conference on Artificial Intelligence*, pp. 119–154, 2004.
- [29] R. Zhou, K. Hwang, and M. Cai, "Gossiptrust for fast reputation aggregation in peer-to-peer networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 9, pp. 1285–1295, 2008.
- [30] L. Xiong and L. Liu, "Peertrust: supporting reputation-based trust for peer-to-peer electronic communities," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 7, pp. 843–857, 2004.
- [31] J. Abawajy, "Determining service trustworthiness in intercloud computing environments," in *Proceedings of the 10th International Symposium on Parallel Architectures, Algorithms, and Networks*, pp. 784–788, 2009.
- [32] S. Marsh, *Formalising trust as a computational concept*. Ph.D thesis, Department of Mathematics and Computer Science, University of Stirling, 1994.
- [33] T. Tran and R. Cohen, "Improving user satisfaction in agent-based electronic marketplaces by reputation modeling and adjustable product quality," in *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 828–835, 2004.
- [34] L. Rasmusson and S. Janssen, "Simulated social control for secure internet commerce," in *Proceedings of the 1996 New Security Paradigms Workshop*, pp. 18–25, 1996.
- [35] *QS University Ranking*. <http://www.topuniversities.com/university-rankings/world-university-rankings/2012/subject-rankings/social-science/communications-media-studies/>, 2012.

- [36] C. D. Luca, "The use of surface electromyography in biomechanics," *Journal of Applied Biomechanics*, vol. 13, no. 135, pp. 135–163, 1997.
- [37] *Buy iTunes Ratings and Comments – Increase iTunes Sales and Downloads*. <http://www.youtube.com/watch?v=CTWV4XaxCo>.
- [38] R. Kerr and R. Cohen, "Smart cheaters do prosper: defeating trust and reputation systems," in *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems*, pp. 993–1000, 2009.
- [39] A. Jøsang and R. Ismail, "The beta reputation system," in *Proceedings of the 15th Bled Electronic Commerce Conference*, pp. 324–337, 2002.
- [40] A. Gelman, *Bayesian data analysis*. Florida, USA: Chapman & Hall/CRC, 2004.
- [41] Y. Wang and M. P. Singh, "Formal trust model for multiagent systems," in *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pp. 1551–1556, 2007.
- [42] A. Jøsang, S. Hird, and E. Facer, "Simulating the effect of reputation systems on e-markets," in *Proceedings of the 1st International Conference on Trust Management (iTrust2003)*, pp. 179–194, 2003.
- [43] J. Weng, C. Miao, and A. Goh, "An entropy-based approach to protecting rating systems from unfair testimonies," *IEICE Trans. INF. and SYST.*, vol. E89-D, pp. 2502–2511, September 2006.
- [44] T. M. Cover and J. A. Thomas, *Elements of information theory*. Wiley, 1991.
- [45] C. Dellarocas, "Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior," in *Proceedings of the 2nd ACM Conference on Electronic Commerce*, pp. 150–157, 2000.
- [46] G. Adomavicius and A. Tuzhilin, "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*, vol. 17, no. 6, pp. 734–749, 2005.
- [47] Z. Noorian, S. Marsh, and M. Fleming, "Multi-layer cognitive filtering by behavioral modeling," in *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems*, pp. 871–878, 2011.
- [48] W. T. L. Teacy, J. Patel, N. R. Jennings, and M. Luck, "TRAVOS: trust and reputation in the context of inaccurate information sources," *Autonomous Agents and Multi-Agent Systems*, vol. 12, no. 2, pp. 183–198, 2006.

-
- [49] B. Yu and M. P. Singh, "An evidential model of distributed reputation management," in *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 294–301, 2002.
  - [50] A. P. Dempster, "A generalization of bayesian inference," *Journal of the Royal Statistical Society, Series B*, vol. 30, pp. 205–247, 1968.
  - [51] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, 1976.
  - [52] M. P. Singh, B. Yu, and M. Venkatraman, "Community-based service location," *Communications of the ACM*, vol. 44, pp. 49–54, 2001.
  - [53] N. Littlestone and M. K. Warmuth, "The weighted majority algorithm," in *Proceedings of the 30th Annual Symposium on Foundations of Computer Science*, pp. 256–261, 1989.
  - [54] B. Yu and M. P. Singh, "Detecting deception in reputation management," in *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 73–80, 2003.
  - [55] D. Heckerman, *A tutorial on learning with Bayesian networks*. Learning in Graphical Models, 1998.
  - [56] W. T. L. Teacy, N. R. Jennings, A. Rogers, and M. Luck, "A hierarchical bayesian trust model based on reputation and group behaviour," in *Proceedings of the 6th European Workshop on Multi-Agent Systems*, pp. 206–212, 2008.
  - [57] R. Jurca and B. Faltings, "Using CHI-scores to reward honest feedback from repeated interactions," in *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 1233–1240, 2006.
  - [58] N. A. Wesis, *Elementary statistics*. Addison-Wesley Publishing Company, 1996.
  - [59] A. Jøsang and J. Haller, "Dirichlet reputation systems," in *Proceedings of the 2nd International Conference on Availability, Reliability and Security (ARES'07)*, pp. 112–119, April 2007.
  - [60] J. Zhang and R. Cohen, "Trusting advice from other buyers in e-marketplaces: the problem of unfair ratings," in *Proceedings of the 8th International Conference on Electronic Commerce (ICEC'06)*, pp. 225–234, 2006.
  - [61] J. Zhang and R. Cohen, "Evaluating the trustworthiness of advice about selling agents in e-marketplaces: a personalized approach," *Electronic Commerce Research and Applications*, vol. 7, no. 3, pp. 330–340, 2008.

- [62] L. Mui, M. Mohtashemi, and A. Halberstadt, “A computational model of trust and reputation,” in *Proceedings the 35th Hawaii International Conference on System Science (HICSS’02)*, pp. 2431–2439, 2002.
- [63] A. Jøsang, “A logic for uncertain probabilities,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 3, pp. 279–311, 2001.
- [64] J. Zhang and R. Cohen, “Combining trust modeling and mechanism design for promoting honesty in e-marketplaces,” *Computational Intelligence*, pp. 1–30, 2012.
- [65] A. Jøsang, R. Ismail, and C. Boyd, “A survey of trust and reputation systems for online service provision,” *Decision Support System*, vol. 43, no. 2, pp. 618–644, 2007.
- [66] N. Miller, P. Resnick, and R. Zeckhauser, “Eliciting informative feedback: the peer-prediction method,” *Management Science*, vol. 51, no. 9, pp. 1359–1373, 2005.
- [67] R. O. Duba, P. E. Hart, and D. G. Stork, *Pattern classification*. Wiley-Interscience, 2001.
- [68] B. Matthews, “Comparison of the predicted and observed secondary structure of t4 phage lysozyme,” *Biochim. Biophys. Acta*, vol. 405, pp. 442–451, 1975.
- [69] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proceedings of the 2nd Interactional Conference on Knowledge Discovery and Data Mining*, pp. 226–231, 1996.
- [70] M. Ankerst, M. M. Breunig, H. P. Kriegel, and J. Sander, “OPTICS: Ordering points to identify the clustering structure,” *ACM SIGMOD Record*, vol. 28, no. 2, pp. 49–60, 1999.
- [71] M. G. Uddin, M. Zulkernine, and S. I. Ahamed, “CAT: a context-aware trust model for open and dynamic systems,” in *Proceedings of the 2008 ACM Symposium on Applied Computing*, pp. 2024–2029, 2008.
- [72] R. Hermoso, H. Billhardt, and S. Ossowski, “Role evolution in open multi-agent systems as an information source for trust,” in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, pp. 217–224, 2010.
- [73] Y. Liu and Y. Sun, “Anomaly detection in feedback-based reputation systems through temporal and correlation analysis,” in *Proceedings of the 2nd IEEE International Conference on Social Computing*, pp. 65–72, August 2010.

-
- [74] B. Yu, M. P. Singh, and K. Sycara, "Developing trust in large-scale peer-to-peer systems," in *Proceedings of the 1st IEEE Symposium on Multi-Agent Security and Survivability*, pp. 1–10, 2004.
- [75] J. Kuzic, J. Fisher, and A. Scollary, "Electronic commerce benefits, challenges and success factors in the australian banking and finance industry," in *Proceedings the 10th European Conference on Information Systems*, pp. 1607–1616, 2002.
- [76] D. Miers, *The strategic challenges of electronic commerce*. <http://www.enix.co.uk/electron.htm>, 2004.
- [77] I. Akyildiz, W. Lee, M. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.
- [78] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.
- [79] S. Mishra, A. Sahai, and R. Brodersen, "Cooperative sensing among cognitive radios," in *Proceedings of IEEE International Conference on Communications*, pp. 1658–1663, 2006.
- [80] C. Dellarocas, "Goodwill hunting: an economically efficient online feedback mechanism for environments with variable product quality," *Agent-Mediated Electronic Commerce IV. Designing Mechanisms and Systems*, vol. LNCD 2531, Springer Verlag, pp. 238–252, 2002.
- [81] R. Jurca and B. Faltings, "An incentive compatible reputation mechanism," in *Proceedings of the 2nd international Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 1024–1027, 2003.
- [82] R. Jurca and B. Faltings, "Towards incentive-compatible reputation management," in *Proceedings of the 2002 International Conference on Trust, Reputation, and Security: Theories and Practice*, pp. 138–147, 2002.
- [83] M. Paolucci, K. Sycara, T. Nishimura, and N. Srinivasan, "Toward a semantic web e-commerce," in *Proceedings of the 6th Conference on Business Information Systems (BIS)*, pp. 153–161, 2003.
- [84] R. Jurca and B. Faltings, "CONFESS: an incentive compatible reputation mechanism for the online hotel booking industry," in *Proceedings of the IEEE Conference on E-Commerce Technology*, pp. 205–212, 2004.

- [85] S. Braynov and T. Sandholm, "Trust revelation in multiagent interaction," in *Proceedings of CHI'02 Workshop on the Philosophy and Design of Socially Adept Technologies*, pp. 57–60, 2002.
- [86] R. K. Dash, S. D. Ramchurn, and N. R. Jennings, "Trust-based mechanism design," in *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems*, pp. 748–755, 2004.
- [87] J. Herlocker, J. A. Konstan, and J. Riedl, "An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms," *Information Retrieval*, vol. 5, no. 4, pp. 287–310, 2002.
- [88] Z. Liang and W. Shi, "Analysis of ratings on trust inference in open environments," *Performance Evaluation*, vol. 65, no. 2, pp. 99–128, 2008.
- [89] F. Li and Y. Yang, "Routing in vehicular ad hoc networks: a survey," *IEEE Vehicular Technology Magazine*, vol. 2, no. 2, pp. 12–22, 2007.
- [90] C. J. Fung, J. Zhang, I. Aib, and R. Boutaba, "Dirichlet-based trust management for effective collaborative intrusion detection networks," *IEEE Transactions on Network and Service Management (TNSM)*, vol. 8, no. 2, pp. 79–91, 2011.