KNOWLEDGE DISCOVERY IN SOCIAL MEDIA: PHYSICAL WORLD, ONLINE WORLD, AND VIRTUAL WORLD

by

Yulei Zhang

A Dissertation Submitted to the Faculty of the

Committee On Business Administration

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY
WITH A MAJOR IN MANAGEMENT

In the Graduate College

THE UNIVERSITY OF ARIZONA

2011

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation

prepared by Yulei Zhang

entitled Knowledge Discovery in Social Media: Physical World, Online World, and Virtual World

and recommend that it be accepted as fulfilling the dissertation requirement for the

Degree of Doctor of Philosophy

____ Date: 04/08/2011
Hsinchun Chen

____ Date: 04/08/2011
Jay F. Nunamaker Jr.

____ Date: 04/08/2011
Susan A. Brown

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

____ Date: 04/08/2011
Dissertation Director: Hsinchun Chen

STATEMENT BY AUTHOR

SIGNED: Yulei Zhang

ACKNOWLEDGMENTS

I want to deeply thank my advisor, Dr. Hsinchun Chen, for his guidance and encouragement throughout my five-year doctoral study at the University of Arizona. Dr. Chen has always been very supportive on my research and my career development. It has been an invaluable opportunity for me to join the Artificial Intelligence Lab under his guidance. I have learned an enormous amount of knowledge from him. Many thanks to my committee members, Dr. Jay F. Nunamaker, Jr. and Dr. Susan A. Brown. Dr. Nunamaker has given me a lot of support and constructive comments on my research. Dr. Brown has always given me detailed suggestions on how to connect theory to technical research. She has also given me a lot of personal support and encouragement. Again, I want to thank all of my dissertation committee members for their guidance and support over the past years. I also want to thank other MIS faculty members for their help. Being in the PhD program at the University of Arizona has been such a precious opportunity for me. I believe it will significantly benefit my future career.

My dissertation has been partly supported by grants from the National Science Foundation (IIS-0428241; CMMI-0738803; IIS-0429364), the Defense Threat Reduction Agency (HDTRA1-09-1-0058), the Library of Congress (LCOSI07C0023), and the United States Department of Agriculture (20063954617579). My work has been supported by many other professors. Many thanks to Dr. Hope Schau in Marketing Department, Dr. Zhu Zhang in MIS department, Dr. Ying Lin in Linguistic Department, and Dr. Paul Hu in OIS department at the University of Utah. All of them have given me a lot of helpful suggestions and encouragement.

I also want to sincerely thank Ms. Cathy Larson for helping me edit my papers and various other types of help and support in the past years. I would also like to thank Li Fan, Shuo Zeng, Chun-Neng Huang, Ximing Yu for working with me on many projects and research studies together. I want to thank all my colleagues and friends, including Chunju Tseng, Hsin-Min Lu, Aaron Sun, Jiesi Cheng, Tianjun Fu, Ping Yan, Yi-Da Chen, David Zimbra, Victor Benjamin, Andy Pressman, and David Ware for their support.

Furthermore, I owe my deepest gratitude to my wife, Yan Dan, my parents, Chao Zhang and Yuling Chen, and the rest of my family. This would not have been possible without all of their support.

DEDICATION

This dissertation is dedicated to my grandma, Huilan Lei, who passed away on July 5th, 2003.

TABLE OF CONTENTS

TABLE OF CONTENTS - *Continued*

TABLE OF CONTENTS - *Continued*

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES - *Continued*

ABSTRACT

Social media have grown tremendously, making the Internet a new platform for community-based social interaction. The rich and vast amount of social media data provides valuable resources for understanding various social phenomena. Different from the world where people physically live, the new media bring additional types of worlds into people's lives: online worlds and virtual worlds. Examples of online worlds include Web forums, blogs, and online reviews, while the most famous example of a virtual world is Second Life. My dissertation is trying to address the overarching questions about how people adapt to social media to share information and exchange opinions, and what factors influence their activities in the new media. I adopt Web mining, machine learning, and computational linguistics techniques to analyze aspects of people and their behavior, such as gender differences, emotional differences, avatar activity differences, and avatar social interaction differences in online and virtual worlds.

Chapter 2 develops a feature-based text classification framework to examine online gender differences between Web forum posters by analyzing writing styles and topics of interest. Guided by the stereotyping and social roles theories, Chapter 3 examines the emotional differences between men and women in text-based online communications. A research framework for automatic emotion detection is developed using sentiment analysis techniques. In the framework, different algorithms are developed to analyze the sentence-level subjectivity and phrase- and word-level polarity. Chapters 4 and 5 focus on investigating avatar behavior in the virtual world. Guided by

the theories of social presence, social role, and gender role, Chapter 4 examines the effects of avatar virtual gender, virtual age, and region theme on avatars' physical activities. Chapter 5 further examines avatars' gender and age differences in their social interactions in help-seeking regions in the virtual world. The overall gender and age difference analyses and detailed investigations by comparing three types of interaction networks based on gender or age are conducted.

Overall, my dissertation contributes to the literature on social media analytics, knowledge discovery, virtual world research, and text and Web mining.

CHAPTER 1. INTRODUCTION

In addition to the world where people physically live, there are two additional types of "worlds" that have become popular. They are online worlds and virtual worlds. Online worlds are based on various Web applications that facilitate interactive information sharing, interoperability, user-centered design, and collaboration on the World Wide Web (http://en.wikipedia.org/wiki/Web_2.0). Examples include Web forums, blogs, online reviews, social-networking sites (e.g., Facebook), and wiki. Virtual worlds take the form of computer-based simulated environments, where users can interact with each other and use and create objects typically in 3D formats (http://en.wikipedia.org/wiki/Virtual_world). The most famous example is Second Life. Other examples include massively multiplayer online games (e.g., World of Warcraft). The vast amount of social media data in these new worlds can provide valuable resources for understanding various social phenomena. To better understand people's adoption of the new worlds, it is important to examine how people communicate in the new worlds, whether they act similarly or differently in the new worlds compared to the physical world, and what factors can influence behavior in the new worlds.

1.1 Online Worlds

The increased use of networked computers in contemporary society has changed the ways in which people communicate with others. Web 2.0 (O'Reilly 2005) has enabled all kinds of communication to take place online. An individual can conduct more interactive communication and act positively using this new medium instead of only passively acquiring information. The two-way communication between end-users and the Web-based communities enhances the information and opinion sharing among Internet users. Examples of online worlds include Web forums, blogs, online reviews, social-networking sites, wiki, chat rooms, instant messages, etc.

Based on the communication synchronicity, online worlds can be classified into two types, synchronous and asynchronous (Guiller and Durndell 2007). Communications in synchronous online worlds, such as chat rooms, take place in real time. In contrast, communications in asynchronous online worlds take place over computers in delayed time (e.g., Web forums, blogs, and wikis). Hence, they allow people to communicate from different places and at different times. The major advantage is that participants can take their time to read and respond to messages (Guiller and Durndell 2007; Harasim 1990).

Understanding aspects of people and their behavior in online worlds could be important for Internet service providers, system developers, information analysts, and end users. Many domains, such as security and marketing, could benefit from such an understanding. The ability of security researchers and analysts to track individual contributors, analyze their trends and views, monitor certain opinion groups, and identify potential threats could be very useful. For the marketing domain, analyzing people's opinions posted in the online worlds can help the sellers adopt and develop services and systems tailored for the different groups of people, and thereby attract more customers.

1.2 Virtual Worlds

The history of scientific and social progress has gone through three waves of change: Agricultural Age, Industrial Age, and Information Age. It is predicted that with the expansion of the Information Age together with the increase in technology advancement, a fourth wave of change will come (Mahmoodi and Jalali 2009). This new age, Virtual Age, is going to make many aspects of people's everyday lives and world affairs to be virtual, such as virtual commerce, virtual learning, and virtual government (Mahmoodi and Jalali 2009). The temporal and spatial limitations in our daily lives will not be an issue in the Virtual Age. The Virtual Age will lead to highly advanced and flexible knowledge-based societies and make them into three-dimensional (3D) virtual worlds.

A virtual world is an electronic environment that mimics the physical world, where people can interact with each other and create virtual objects (Bainbridge 2007). Second Life (http://secondlife.com/) is currently one of the most popular 3D virtual worlds. Second Life residents interact with each other through their own customized avatars. Avatars can explore different regions, socialize with other avatars, participate in individual and group activities, and create and trade virtual property and services (http://en.wikipedia.org/wiki/Second_Life). In Dec 2009, there are 769 thousand active users in Second Life (http://blogs.secondlife.com/).

It is expected that virtual worlds will grow in societal importance and will influence various aspects of people's lives (Messinger et al. 2009). Virtual worlds can also help researchers and scientists expand the scope of their research (Bainbridge 2007).

They can enable experiments to be scaled up to hundreds or thousands of subjects, include subjects from underrepresented groups, and conduct experiments that need longer periods of time (e.g., months) in the physical world.

In addition, a virtual world serves as a new place to do business effectively in new ways. It is important to understand the marketing strategies related to advertising, retailing, and eCommerce as well as customer relationships in this new platform (Hemp 2006; Holzwarth et al. 2006). For example, Holzwarth et al. (2006) found that using avatar sales agents on eCommerce websites could help increase user satisfaction, and led to a more positive attitude toward the product.

1.3 Dissertation Framework

As shown in Figure 1.1, my dissertation framework contains three key elements: physical world, online world, and virtual world.

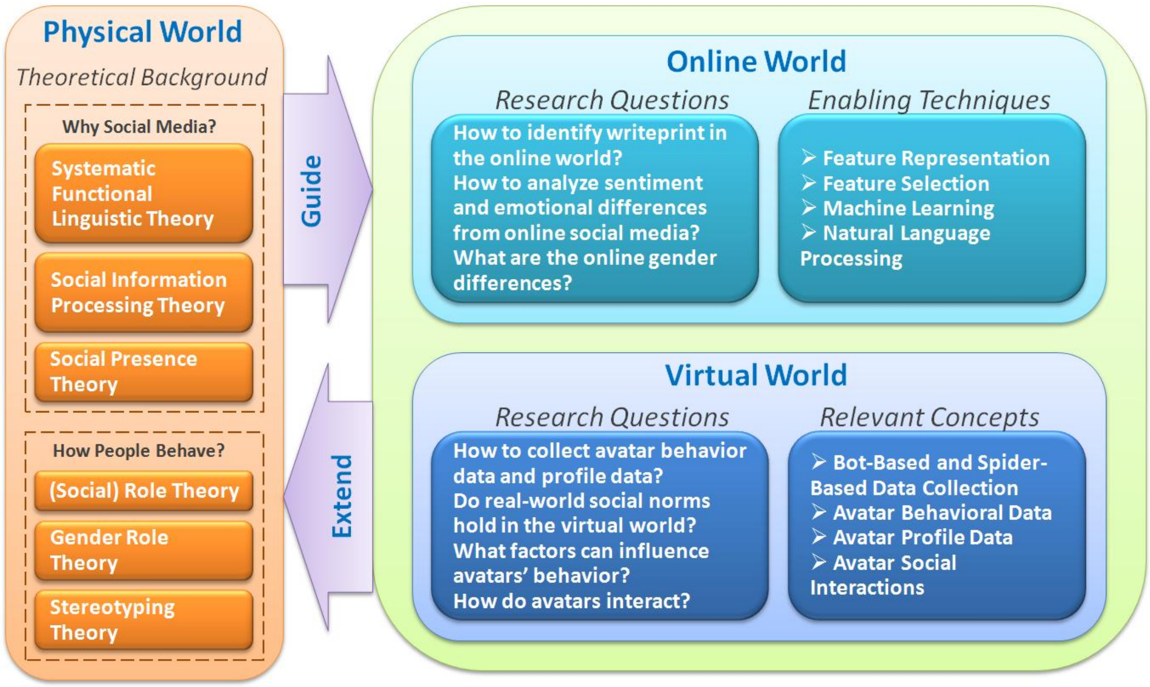

Figure 1.1 My dissertation framework

Theories developed in the physical world are used to guide the understanding and analysis of people's behavior in the online and virtual worlds. Related theories include the Systematic Functional Linguistic Theory (Halliday 1985; Halliday and Matthiessen 2004), Social Information Processing Theory (Walther 1992), Social Presence Theory (Short et al. 1976), (Social) Role Theory (Biddle 1986), Gender Role Theory (Eagly and Karau 1991), and Stereotyping Theory (Brebner 2003; Brody 1997).

The first three theories can be leveraged to examine the effectiveness of social media communications in the online and virtual worlds. The Systematic Functional Linguistic Theory (Halliday 1985; Halliday and Matthiessen 2004) states that language is a social semiotic resource that people can utilize to express meanings in a given context.

It places the function of language as central, and suggests that language has three types of meaning including ideational, textual, and interpersonal. Therefore, language has style and structure, contains ideas, and can be used for people to communicate with each other (Abbasi et al. 2008a). According to Social Information Processing Theory (Walther 1992), it takes longer to develop an online interpersonal relationship than a traditional face-to-face relationship. However, once established, personal relationships in the online world can have the same level of qualities as face to face relationships in the physical world. Social Presence Theory (Short et al. 1976) classifies different communication media based on the social presence degree which refers to the level of awareness of the other person in a communication interaction. Using communication medium with appropriate social presence can lead to effective communication.

The other three theories can be leveraged to guide the understanding of people's behavior in the online and virtual worlds. (Social) Role Theory (Biddle 1986) states that people act according to certain socially defined roles that contains a set of rights, duties, expectations and norms. According to the theory, an individual's behavior is context specific and based on his/her social position and other related factors. Therefore, in general, people behave in a predictable way. Based on gender related social roles, Gender Role Theory (Eagly and Karau 1991) states that males and females behave differently since cultural expectations of the two genders are different. Each gender identity has different sets of roles considered as socially appropriate. According to Stereotyping Theory (Brebner 2003; Brody 1997), a stereotype is a common belief about a particular social group. In most cases, the stereotypes of a social group are subjective rather than objective. Gender related stereotypes have been formed historically with occupation differences of the two genders.

Research questions related to the online world include:

- How to identify writeprint in the online world?
- How to analyze sentiment and emotional differences from online social media?
- What are the online gender differences?

Techniques that can be used to address the above research questions include: feature representation, feature selection, machine learning, and natural language processing (NLP). Feature representation is used to transform full text documents to document vectors (called features) which describe the main contents of the documents. Compared with the original documents, features can be more easily handled by machine learning algorithms. Different types of features such as lexical features, syntactic features, structural features, and content-specific features can be used to represent the documents (Zheng et al. 2006). Lexical features refer to character- or word-based statistical measures of lexical variation; syntactic features refer to the patterns used to form sentences; structural features show the text organization and layout; and content-specific features are comprised of important keywords and phrases on certain topics. No matter which feature representation is used, a typical textual dataset usually has a large number of features. However, not all the features are necessary. Many of them may be noisy or redundant. Therefore feature selection that aims at identifying a minimal-sized subset of features relevant to the target concept can be applied (Dash and Liu 1997). The objective of feature selection is threefold: improving the prediction accuracy, providing faster and more cost-effective prediction, and providing a better understanding of the underlying process that generated the data (Li et al. 2007). Machine learning and Natural Language Processing (NLP) techniques (such as classification algorithms) can be used to further investigate patterns and conduction prediction based on the textual data.

Research questions related to the virtual world include:

- How to collect avatar behavioral data and profile data?
- Do real-world social norms hold in the virtual world?
- What factors can influence avatars' behavior?
- How do avatars interact?

To answer the above research questions, the first important step is to examine how to conduct data collection in the virtual world. The bot-based approach can be used to collect avatars' behavioral data. Bots can be sent to the virtual world regions as real avatars and automatically collect other avatars' behavioral data. By using LibOpenMetaverse, one bot is able to cover the entire area of one region and no human intervention is needed to control the bot. However, the limitation of the bot-based approach is that it cannot collect the avatar profile data. To collect such information, the spider-based approach can be adopted. Automatic spidering programs can be developed to collect avatar profile pages in HTML format. Parsing programs can then be created to parse out the data fields of avatar identity information such as name, birthday, profile picture, etc. By combining both the bot-based approach and the spider-based approach into one framework, both avatar behavioral and profile data can be collected. In addition to examining individual avatar’s activities, with the position and timestamp information collected, it is also possible to create avatar interaction networks and further investigate avatars’ social interaction patterns.

1.4 Summary of Dissertation Chapters

Chapter 2 studies gender differences in online world communications. Specifically, feature-based online social media text classification techniques are used to investigate online gender differences between female and male participants in Web forums, by examining their writing styles and topics of interest. The feature-based gender classification framework is generic and can be applied to different Web forums. In the framework, different types of features are examined, including lexical features, syntactic features, structural features, and content-specific features. For content-specific features, unigrams and bi-grams are automatically extracted from the entire forum instead of manually selected. Five feature sets are developed by adding content-specific features to the basic content-free features and conducting feature selection. The experimental results showed that the feature sets combining both content-specific and content-free features performed significantly better than the ones consisting of only content-free features. In addition, feature selection on large feature sets improved the classification performance significantly. The results also indicated the existence of online gender differences in Web forums. Through further investigation on the selected feature set, different topics of interest between females and males were identified.

In terms of emotional differences, previous studies have found that in physical world communications, women tend to be more emotional than men, and they are also more likely to express both positive emotions (such as happiness and love) and negative emotions (such as fear and sadness) than men. Guided by stereotyping and social roles theories, Chapter 3 further examines the emotional differences between women and men in online world communications by developing an automatic emotion detection framework based on sentiment analysis techniques. Different algorithms are developed and incorporated in the framework, including sentence-level, phrase-level, and word-level emotion detection algorithms. The experimental results on a large and long-standing international women's political forum showed that women were more likely to express their opinions subjectively than men at the sentence-level, and they were more likely to express both positive and negative emotions at the phrase-level as well as at the word-level.

With the increased popularity of virtual worlds, nowadays hundreds of thousands of people from different physical locations can join virtual worlds whenever they want to interact with each other and create objects in a computer-based simulated environment, typically in an interactive 3D format. The rich social media data generated in virtual worlds has important implications for business, education, social science, and society at large. In order to fully use the benefits of virtual worlds, it is important to know whether people display behavior that is consistent between the virtual world and the physical world. However, the best way to collect avatar-related data from virtual worlds is a relatively unexplored topic and remains an issue. As an exploratory study, Chapter 4 investigates whether real-life social norms hold in the virtual world and what the major factors are that can influence people's behavior in the virtual world. Guided by the theories of social presence, social role, and gender role, the main effects and interaction effects of avatar virtual gender, virtual age, and region theme on avatars' physical activities are examined. To enable the analyses, an integrated technical framework for avatar data collection is developed by combining an improved bot-based approach and a spider-based approach. The experimental results indicated that, in general, male avatars were more (less) likely to perform high-active (low-active) actions than female avatars; young-aged avatars were more (less) likely to perform high-active (low-active) actions than old-aged avatars; avatars in commercial transaction regions were more (less) likely to perform high-active (low-active) actions than avatars in help-supporting regions.

In addition to understanding individual avatar behavior, Chapter 5 further explores avatars' social interaction patterns when seeking help in the virtual world. The technical framework developed in Chapter 4 is used for data collection. Degree, betweenness centrality, HITS, and PageRank scores are used to measure avatar interactions. Both the overall avatar gender and age difference analyses and detailed investigations are conducted. In the detailed analyses, three types of interaction networks based on gender or age are compared, including the male-male, female-female, and male-female/female-male interaction networks for gender and the young-young, old-old, and young-old/old-young interaction networks for age. The overall analysis results showed that the old-aged avatars had larger degree and higher betweenness centrality, HITS authority/hub, and RageRank scores than young-aged avatars. This may indicate that when avatars had problems, they were more likely to interact with old-aged avatars to seek help, and old-aged avatars tended to be in the more important positions in the network. The detailed analysis on avatar gender difference indicated that female avatars had more interactions with female avatars when seeking help compared with interactions among male avatars. In addition, both male-male and female-female interaction networks had significantly higher betweenness centrality, HITS, and PageRank scores than the male-female/female-male interaction network, indicating that the interaction networks of the same gender tended to be more centralized than the network of interactions between the two genders. The detailed analysis on avatar age difference showed that the old-old interaction network had the largest degree, followed by the young-old/old-young interaction network and the young-young interaction network, suggesting that when seeking help, both old-aged and young-aged avatars were more likely to interact with old-aged avatars. In addition, both young-young and old-old interaction networks had significantly higher betweenness centrality, HITS, and PageRank scores than the young-old/old-young interaction network, indicating that on average the interaction networks of the same age group tended to be more centralized than the interaction network of the two different age groups.

Chapter 6 concludes my dissertation. It highlights the major research contributions, relevance to MIS research, and future research directions.

CHAPTER 2. GENDER CLASSIFICATION FOR WEB FORUMS

2.1 Introduction

The rapid development and evolution of the Internet has enabled people to access information whenever and wherever they want. Recently, with the advent of Web 2.0, the Internet has evolved towards multimedia-rich content delivery, end-user content generation, and community-based social interaction (O'Reilly 2005). More and more Web forums, blogs, wikis and other social media have been generated and become extremely popular. Such Web 2.0 social media help enhance information sharing, opinion generation, and community-based discussion for various emerging social and political topics.

Although it has a male-dominated history, the Internet is increasingly becoming a new medium for women to share their concerns and express opinions about personal, social and political issues (Harcourt 2000). Women could gain equal presence or influence with men in the virtual community. In addition, their desire for gender equality continues to influence their Internet contributions and writings. Meanwhile, the increasing availability of the Internet offers marginalized groups and individuals a voice in the public sphere (Harp and Tremayne 2006; Mitra 2004). For example, Harcourt (2000) mentions the increasing voice of local Arab women on a global level through the Internet; Mitra (2004) argues that the Internet has allowed women in South Asia be heard by the outside world.

In many disciplines, questions concerning gender differences in the context of online communication have been raised (Halbert 2004). Online gender differences (i.e., the digital gender gap in some studies), which refers to the differences between women and men in Internet use, have been shown and studied in previous research (Fountain 2000; Fuller 2004; Harp and Tremayne 2006). Some studies point out that women are less likely to express political opinions and tend to have a less authoritative manner in their conversation style (Ogan et al. 2005). More research is critically needed to explain online gender differences in social, political, and even business (e.g., online shopping) activities.

Understanding online gender differences and why they occur could be important for Internet service providers, system developers, information analysts, and end users. Many domains, such as security and marketing, could benefit from such an understanding. The ability of security researchers and analysts to track individual contributors, analyze gender-specific trends and views, monitor certain opinion groups, and identify potential threats could be very useful. For the marketing domain, a better understanding of the different interests in various products between the two genders can help the sellers adopt and develop services and systems tailored for the two groups of people, and thereby attract more customers.

In this study, I adopted feature-based text classification techniques to identify and analyze online gender differences by examining the discrepancy between women's and men's writing styles. Most previous online text classification studies have focused on authorship classification and sentiment classification; relatively less effort has been put on gender classification. To improve classification performance, the most recent authorship and sentiment classification studies incorporated all four types of features including lexical, syntactic, structural, and content-specific features. However, for gender classification, no existing study has used all four types of features. As to the context, previous gender classification studies have mainly focused on novels (Hota et al. 2006), non-fiction articles (Koppel et al. 2002), e-mails (Corney et al. 2002), and Web blogs (Nowson and Oberlander 2006; Schler et al. 2006). Relatively less effort has been put on the Web forum context. In addition, even for those focusing on Web forum context, most studies used basic keyword-based analysis with a relatively small set of keywords to examine the topic differences between males and females (Guiller and Durndell 2007; Seale et al. 2006). Few studies have investigated gender differences in Web forums using feature-based text classification techniques. Therefore, I proposed a feature-based gender classification framework to analyze online gender differences for Web forums by examining the writing styles and content (including different types of linguistic features) of female and male posters. The experiment was conducted on an Islamic women's political forum, and I compared the performance of different feature sets. The best classification results were achieved by incorporating all four types of features and conducting feature selection, demonstrating the efficacy of this framework for gender classification for Web forums. I further analyzed the different topics preferred by women and men respectively.

The remainder of this chapter is organized as follows. Section 2.2 provides a review of previous research in online gender differences and online text classification.

Section 2.3 describes research motivation, while Section 2.4 presents the research design. Section 2.5 describes the experiment used to evaluate the effectiveness of the proposed framework and discusses the results. Section 2.6 concludes the paper with closing remarks and future directions.

2.2 Literature Review

2.2.1 Online Gender Differences

With the increasing availability and popularity of the Internet, as well as the advent of Web 2.0, more and more women participate in community-based social media (Consaluo and Paasonen 2002). The Internet, therefore, has become a medium for women to share their political opinions and knowledge (Harcourt 2000). They are also creating their own online networks to exchange information and ideas (Sherman 2001).

The Internet is not only useful as a fast communication medium, it is also a very crucial channel of information on women's rights issues. Women use the Internet to fight against violence by building a strong layer of support through which their personal struggles can be discussed and solutions shared (Harcourt 2000). As an example, Harcourt (2000) talks about a case regarding a Muslim woman's right of choice of marriage in her study; she argues that "we could, within hours, receive case law on the issue from other Muslim countries as well as legal and scholarly opinions and references that prove critical in winning the case."

Researchers have also shown an increasing interest in studying online gender differences, which refers to the fact that there exist differences between women and men in Internet use (Bimber 2000). Previously, the major online gender difference noted was that fewer women than men used the Internet. For example, the A. C. Nielsen CommerceNet consortium from 1999 showed that among U.S. and Canadian Internet users, 53% were men and 47% were women; among online shoppers, 62% were men and 38% were women; and among people who reported having used the Internet in the last twenty-four hours for any purpose, 68% were men and 32% were women (CommerceNet 1999). In the realm of political activity, the National Election Studies (NES) data showed that visitors to Internet campaign sites during the 1998 election season were 60% male and 40% female (National Election Study 1998). However, with the rapid development and increasing availability of the Internet, more and more women are accessing the Internet to acquire information, express their ideas, and share common concerns. The May 2008 survey by the Pew Internet and American Life Project found that 73% of men and 73% of women use the Internet (Pew Internet and American Life Project 2008). In contrast its 2004 survey reported 66% and 61% Internet use for men and women respectively.

Although access to technology is not an issue today, women and men do have differences in Internet use depending on motivation and interest in the content being produced and consumed (Harp and Tremayne 2006). Jackson et al. (2001) found that women are more likely to use the Internet as a communication tool and men are more likely to use it as a means of information seeking. According to Ogan et al. (2005), women are less likely to express political opinions and tend to have a less authoritative manner in their conversation style. Meanwhile, some studies (Fuller 2004; Youngs 2004) observed that women's concerns tend to center around the private sphere of life; i.e., the domestic sphere of home, family, private relations, and sexual reproduction; on the other hand, men are more externally focused on the public sphere and political realm including government and commercial establishments.

As to online communication in Web forums, previous studies have used keyword analysis to show that women and men do have different topics that they are interested in and care about (Seale et al. 2006). Seale et al. (2006) analyzed cancer-related Web forums and found that women's discussions are more likely to lean towards the exchange of emotional support, including concern with the impact of illness on a wide range of other people; however, men are more likely to participate in threads related to treatment information, medical personnel, and procedures. Guiller and Durndell (2007) analyzed an online course discussion board and found that women are more likely to explicitly agree and support others and make more personal and emotional contributions than men; on the other hand, men are more likely to use authoritative language and to respond negatively in interactions than women.

2.2.2 Online Text Classification

In this study, I adopted online text classification techniques to study the online gender differences in Web forums by examining the writing style of the posted messages.

Online text classification has several important characteristics, including various types of problems, features, and online texts.

2.2.2.1 Different Types of Online Text Classification Problems

With the advent of Web 2.0, more and more automatic classification studies using online text-based social media data have appeared. In those studies, the investigated classification problems mainly include authorship, sentiment and gender classification. Unlike the classical topic-based classification problem in information retrieval, social media classification relies heavily on the information and writing styles of authors in various online social media.

Authorship classification aims to determine which author produced which piece of writing by examining the styles and contents of writings produced by different authors. Previous studies have applied authorship classification to various online social media texts. De Vel and his collaborators (deVel 2000; deVel et al. 2001) applied the conventional text classification methods to identify the authors of e-mails. A recent comprehensive study conducted by Abbasi and Chen (2008) tested their newly developed Writeprints technique with a rich set of features on various online datasets, including e-mails, instant messages, feedback comments, and program codes.

Sentiment classification for online texts aims to analyze direction-based texts (i.e., texts containing opinions and emotions) to determine whether a text is objective or subjective, or whether a subjective text contains positive or negative sentiments. The common, two-class sentiment classification problem involves classifying sentiments as positive or negative (Pang et al. 2002; Turney 2002). However, additional variations include classifying sentiments as opinionated/subjective or factual/objective (Wiebe et al. 2001; Wiebe et al. 2004). Instead of sentiments, other studies have attempted to classify emotions, including happiness, sadness, anger, horror, etc. (Grefenstette et al. 2004; Subasic and Huettner 2001).

Gender classification aims to determine whether a piece of writing was produced by a female or male by examining the writing styles and contents of female and male authors. Gender classification is different from authorship classification in that authorship classification examines individual differences of people’s writing styles no matter whether a person is a woman or a man, while gender classification is used to examine and identify the overall differences in writing styles between the two gender groups, in order to gain an understanding of gender-based differences. Previous gender classification studies using automatic text classification techniques have been done on both traditional articles (e.g., novels and non-fiction articles) and online social media texts (e.g., e-mails and Web blogs). As an example of gender classification on traditional articles, Koppel et al. (2002) used the Exponential Gradient (EG) algorithm to classify genders for both fiction and non-fiction documents. By using a feature set combining function words and parts-of-speech (POS) tags, they achieved 79.5% accuracy for fiction documents and 82.6% accuracy for non-fiction documents. After feature selection, the accuracy increased to 98% for both fiction and non-fiction documents. Another study conducted by Hota et al. (2006) classified the gender of Shakespeare's characters based on a collection of his plays. They achieved the highest accuracy of 74.28% using Support Vector

Machine (SVM) on the feature set consisting of both content-independent and content-based features. Argamon and his collaborators (2003a) analyzed writing styles and identified a set of lexical and syntactic features that differed significantly according to author gender in both fiction and nonfiction documents. In particular, they found that although the total number of nominals used by female and male authors was virtually identical, females used many more pronouns and males used many more noun specifiers.

For online social media text, most previous gender classification studies focused on e-mails (Corney et al. 2002) and Web blogs (Nowson and Oberlander 2006; Schler et al. 2006). Corney et al. (2002) used SVM to classify genders for e-mails and achieve the highest F-measure of 71.1% using a combination of lexical features, structural features and selected gender-specific features. Nowson and Oberlander (2006) used SVM to classify genders for Web blogs and achieved the highest accuracy of 91.5% using a combination of parts-of-speech (POS), bi-grams, and trigrams as the features. Schler et al. (2006) also conducted gender classification on Web blogs and emphasized the significant differences in writing styles and contents between female and male bloggers as well as among authors of different ages.

2.2.2.2 Features for Online Social Media Text Classification

Features are very important for online social media text classification. Good feature sets can improve the performance of the classifier. There are four types of features that were often used in previous online social media text classification studies:

lexical, syntactic, structural, and content-specific features. Among them, the first three types are content-free features; the fourth type contains features related to specific topics.

Lexical features refer to character- or word-based statistical measures of lexical variation. Lexical features mainly include: character-based lexical features (Argamon et al. 2003b; Gamon 2004), vocabulary richness measures (Yule 1944), and word-based lexical features (deVel et al. 2001; Zheng et al. 2006). Examples of character-based lexical features include the total number of characters, the number of characters per sentence, the number of characters per word and the usage frequency of individual letters. Examples of vocabulary richness measures include the number of words that occur once and twice, and some other statistical measures defined by Yule (1944). Examples of word-based lexical features include the total number of words, the number of words per sentence, and word length distribution.

Syntactic features refer to the patterns used to form sentences. Commonly studied syntactic features are function words (Koppel et al. 2006; Koppel et al. 2002), punctuation (Baayen et al. 2002), and part-of-speech (POS) tags (Argamon et al. 1988; Baayen et al. 2002; Gamon 2004; Nowson and Oberlander 2006). These studies also demonstrated that syntactic features may be more reliable compared with lexical features. To study the writing style differences between females and males, Argamon and his collaborators (2003a) used over 1,000 features including 467 function words and a set of POS tags.

Structural features show the text organization and layout. They are especially useful when studying online social media texts (deVel et al. 2001). Traditional structural features include greetings, signatures, the number of paragraphs and the average paragraph length (deVel et al. 2001; Zheng et al. 2006). Other structural features include technical features such as the use of various file extensions, font sizes, and font colors (Abbasi and Chen 2005).

Different from the above content-free features, content-specific features are comprised of important keywords and phrases on certain topics (Martindale and Mckenzie 1995; Zheng et al. 2006) such as word n-grams (Abbasi and Chen 2005; Diederich et al. 2003; Nowson and Oberlander 2006). Usually, these features represent specific subject matter in a given domain. For example, content-specific features on a discussion of computers may include “laptop” and “notebook.” Previous studies have showed that content-specific features can improve the performance of online text classification (Abbasi and Chen 2005; Abbasi et al. 2008b; Schler et al. 2006; Zheng et al. 2006).

2.2.2.3 Different Types of Online Social Media Texts

The major types of online social media texts include e-mails, online news, Web blogs, online reviews, and Web forums. Among them, e-mail and online news typically belong to Web 1.0, while Web blogs, Web forums, and online reviews are considered to be Web 2.0 media.

Different from the other types of online social media texts, e-mails can only be used to share information between the senders and receivers. Typically, the general public cannot access the content. In contrast, online news is always available for any

Internet user. However, it enables only a one-way flow of information through static websites that contain “read-only” materials. Therefore, users can only passively acquire information instead of actively participating in the discussions.

As Web 2.0 media, Web blogs, online reviews, and Web forums contain a great deal of dynamic user-generated content. Different people can participate in and exchange opinions through these communication platforms. For Web blogs, the blog owner typically leads the discussion and others can follow up with their comments. Compared to Web blogs, online reviews and Web forums tend to have relatively more “balanced” discussions among participants. “Balanced” here refers to the number of participants and the number of discussion messages they posted. In forums, participants are generally free to initiate their own discussions on topics of their own choosing, as opposed to in blogs, topics are generally set by the blog owner. One major difference between online reviews and Web forums is that online reviews are more focused on a particular product or category of products, while the discussion topics in Web forums tend to be broader, meaning that in addition to certain products, people also share their opinions on certain events or social and political issues.

Previous gender classification studies have mainly focused on either traditional articles (e.g., novels and non-fiction articles) or e-mails (Corney et al. 2002) and Web blogs (Nowson and Oberlander 2006; Schler et al. 2006). Relatively less effort has been put on the Web forum context. Considering its role as a major type of online social media with a balanced nature of discussions among participants and a relatively broader range of topics, it is important to understand the online gender differences in Web forums.

2.2.2.4 Feature Selection for Text Classification

To perform text classification, a textual dataset is usually represented by a set of features. When the number of features is large, not all the features are necessary to learning the concept of interest; instead, many of them may be noisy or redundant; feeding all these features into a model often results in over fitting and poor predictions (Meiri and Zahavi 2006). In such cases, feature selection can be used to improve the classification performance by selecting an optimal subset of features (Dash and Liu 1997; Guo and Nixon 2009). Previous text classification studies using n-gram features usually included some form of feature selection in order to extract the most important words or phrases (Koppel and Schler 2003). The objective of feature selection is threefold: to improve the prediction accuracy, provide faster and more cost-effective prediction, and provide a better understanding of the underlying process that generated the data (Hu et al. 2007; Li et al. 2007). A feature selection method generates different candidates from the feature space and assesses them based on some evaluation criterion to find the best feature subset (Li et al. 2007). Graph-based search algorithms are often used to find the optimal features (Li et al. 2007). In general, when the feature set is large, using feature selection in text classification can improve the classification performance by offering more concise and precise feature representations of documents (Scott and Matwin 1999).

2.3 Motivation

Understanding online gender differences and why they occur could be important for Internet service providers, system developers, information analysts, and end users. Many domains, such as security and marketing, could benefit from understanding gender-based differences. The ability of security researchers and analysts to track individual contributors, analyze gender-specific trends and views, monitor certain opinion groups, and identify potential threats could be very useful. For the marketing domain, a better understanding of the different interests in various products between the two genders could help sellers adopt and develop services and systems tailored for the two groups of people, and thereby attract more customers.

According to the literature review, most previous online text classification studies focused on authorship classification and sentiment classification; relatively less effort has been put on gender classification. To improve the classification performance, the most recent authorship classification studies (Abbasi and Chen 2005; Zheng et al. 2006) have incorporated all four types of features, i.e., lexical, syntactic, structural, and content-specific features. Although early sentiment classification studies often used one type of feature, later studies (Gamon 2004; Wiebe et al. 2004) added other types of features to improve the classification performance. However, for gender classification, no existing study has been seen to use all four types of features. Although previous studies have shown the existence and evolution of online gender differences and the importance of gender role in political movements, most of them have focused on either traditional articles (e.g., novels and non-fiction articles) or e-mails (Corney et al. 2002) and Web blogs (Nowson and Oberlander 2006; Schler et al. 2006). For the relatively few studies in the Web forum context, most of them used basic keyword-based analysis (Guiller and Durndell 2007; Seale et al. 2006) instead of the more advanced feature-based text classification techniques.

Motivated by the above discussion, in this study I proposed a framework to investigate online gender differences in the context of Web forums using feature-based gender classification techniques by incorporating all four types of features. I intended to answer the following research questions raised from the literature review.

1. Can gender classification techniques be used to identify and analyze online gender differences in Web forums?

2. Will the use of both content-free features (i.e., lexical, syntactic, and structural features) and content-specific features improve gender classification performance for Web forums compared to using only content-free features?

3. For relatively large feature sets, will feature selection that returns a smaller number of the most important features improve the gender classification performance for Web forums?

The first question is more general while the other two are more specific. Finding the answer to the first question is the primary motivation for the study as a whole. The efficacy of the proposed gender classification framework is addressed later in this paper. In order to answer the remaining two questions, I developed the detailed hypotheses listed in Section 2.5.2, which in turn drove the details of the study design.

2.4 Research Design

In order to address these questions, I developed a framework of feature-based gender classification on Web forums. As shown in Figure 2.1, the framework includes several essential components: Web forum message acquisition, feature generation, and classification and evaluation.

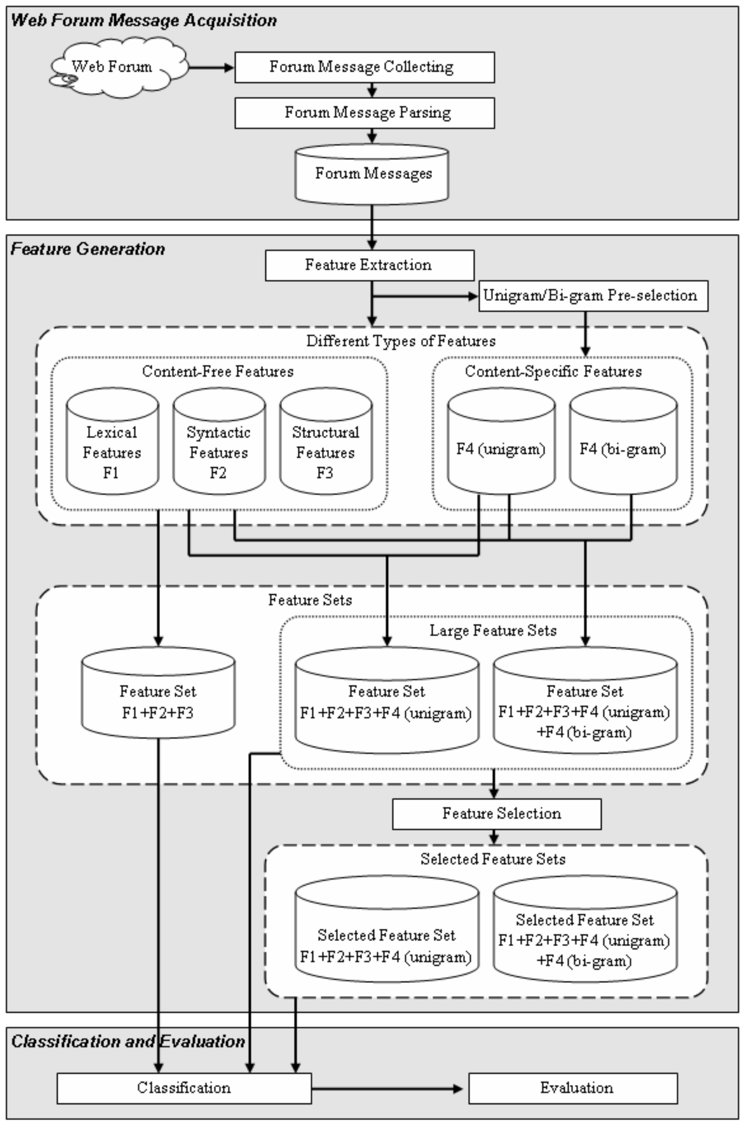

Figure 2.1 Research framework

2.4.1 Web Forum Message Acquisition

This component consisted of two steps: forum message collecting and forum message parsing. First, spidering programs were developed to collect all the messages in a given open source Web forum as HTML pages. After that, parsers were built to parse out the message information from the raw HTML pages and store the parsed data in a relational database.

2.4.2 Feature Generation

In this component, different feature sets containing different types of features were generated. By doing so, I could compare and evaluate the performance of different feature sets in gender classification for Web forums in order to answer research questions 2 and 3.

There were several steps in this component: feature extraction, unigram/bi-gram pre-selection, and feature selection. Each of these steps led to the generation of different feature sets.

Feature Extraction: Different types of features were extracted based on all messages collected from a given open source Web forum. In this study, I extracted the lexical (denoted by F1), syntactic (denoted by F2), and structural (denoted by F3) features as content-free features, and unigrams (denoted by F4(unigram)) and bi-grams (denoted by F4(bi-gram)) as content-specific features.

For F1 features, the character-based lexical features used (2000; 1996; 1994); the vocabulary richness features (1998); and the word-length frequency features (2001; 1887) were adopted. In total, 87 lexical features were used. For F2 features, a set of 150 function words used in Zheng et al. (2006) were adopted, since this study also focused on Web forum messages, although it is about authorship classification. In addition, 8 punctuation marks suggested by Baayen et al. (1996) were adopted. Therefore, a total of 158 syntactic features were used. For F3 features, I chose five of the most common ones from previous literature (Abbasi and Chen 2005; 2006) that could be applied to a broad number of general Web forums: the total number of sentences per message, the total number of paragraphs per message, the number of sentences per paragraph in a message, the number of characters per paragraph in a message, and the number of words per paragraph in a message. I did not use a large number of structural features related to technical structures (e.g., font colors and font sizes), because some Web forums may not have had the related characteristics. For example, some popular (but old) Web forums do not have functions which allow users to change the font colors and font sizes.

Unigram/Bi-gram Pre-selection: Although content-free features are important for online text classification, content-specific features that consist of important keywords and phrases on certain topics could be more meaningful, thus leading to relatively high representative ability. Content-specific features used in previous online text classification studies are either a relatively small number of manually selected, domain-specific keywords (Li et al. 2006; Zheng et al. 2006), or a relatively large number of n-grams automatically learned from the textual data collection (Abbasi and Chen 2005; Abbasi et al. 2008b; Peng et al. 2003; Schler et al. 2006). The large potential feature spaces of n-grams have been shown to be effective for online text classification (Abbasi and Chen 2008). Therefore, this study used n-grams as content-specific features. Specifically, unigrams (i.e., F4(unigram)) and bi-grams (i.e., F4(bi-gram)) were used. The unigrams and bi-grams were extracted from all the messages in the Web forum. After removing the stop-words, I kept the unigrams and bi-grams that appeared more than ten times in the whole forum as the content-specific features.

By conducting feature extraction and unigram/bi-gram pre-selection, five types of features were obtained. Based on those different types of features, three feature sets were built in an incremental way: (1) feature set F1+F2+F3, (2) feature set F1+F2+F3+F4(unigram), and (3) feature set F1+F2+F3+F4(unigram)+F4(bi-gram). This incremental order represents the evolutionary sequence of features used for online text classification (Abbasi and Chen 2008; Zheng et al. 2006). Studies (Abbasi and Chen 2008; Zheng et al. 2006) have shown that lexical and syntactic features are the foundation for structural and content-specific features. In this study, feature set F1+F2+F3 was used as the baseline feature set to assess the performance of the other two proposed feature sets which also incorporates content-specific features.

Feature Selection: By adding unigrams and unigrams plus bi-grams as content-specific features respectively, feature sets F1+F2+F3+F4(unigram) and F1+F2+F3+F4(unigram)+F4(bi-gram) can be very large. Feature selection was conducted on them using the Information Gain (IG) heuristic due to its reported effectiveness in previous online text classification research (Abbasi and Chen 2008; Koppel and Schler

2003), thus building the selected feature sets F1+F2+F3+F4(unigram) and F1+F2+F3+F4(unigram)+F4(bi-gram). In addition to IG, there are some other feature selection methods that have been reported to have effective performance. For example, Forman (Forman 2003) found that in general Bi-Normal Separation (BNS) and IG are the two most effective methods compared to other feature selection methods. BNS sometimes performed "marginally" better than IG. However, the gap was barely visible in some cases. In terms of precision, "Information Gain yielded the best result most often" (Forman 2003).

As defined in the following formula, Information Gain IG(C,A) measures the amount of entropy decrease on a class C when providing a feature A (Quinlan 1986; Shannon 1948). The decreasing amount of entropy reflects the additional information gained by adding feature A. In the formula, H(C) and H(C|A) represent the entropies of class C before and after observing feature A respectively. The Information Gain for each feature varies along the range 0-1 with higher values indicating more Information Gained by providing certain features.

$IG(C, A) = H(C) - H(C \mid A)$, where:

$$H(C) = -\sum_{c \in C} p(c) \log_2 p(c),$$
$$H(C \mid A) = -\sum_{a \in A} p(a) \sum_{c \in C} p(c \mid a) \log_2 p(c \mid a).$$

All features with an information gain greater than 0.0025 (i.e., IG(C,A) > 0.0025) are selected (Abbasi et al. 2008b; Yang and Pedersen 1997). The idea is to try to achieve the best classification performance by filtering out the features with less to contribution while keeping the ones with relatively higher discriminatory powers.

2.4.3 Classification and Evaluation

To assess the performance of each feature set, the standard classification performance metrics were adopted - i.e., accuracy, precision, recall, and F-measure. These metrics have been widely used in information retrieval and text classification studies (Abbasi and Chen 2008; Abbasi et al. 2008b; Li et al. 2008), especially for datasets with balanced data points among different classes. In this study, the testbed was quite balanced between the two gender groups, so I chose to use these performance measures. When data sets are unbalanced, the ROC curve can be used as another important performance measure (Seiffert et al. 2010). SVM was used as the classifier because of its often reported best performance in many previous online text classification studies (Abbasi and Chen 2008; Abbasi et al. 2008b; Hu et al. 2007; Li et al. 2006; Zheng et al. 2006).

Among the four standard measures, accuracy assesses the overall classification correctness; while precision, recall, and F-measure evaluate the correctness of each class:

$$\text{accuracy} = \frac{\text{number of all correctly classified Web forum messages}}{\text{total number of Web forum messages}},$$

$$\text{precision (i)} = \frac{\text{number of correctly classified Web forum messages for class i}}{\text{total number of Web forum messages classified as class i}},$$

$$\text{recall (i)} = \frac{\text{number of correctly classified Web forum messages for class i}}{\text{total number of Web forum messages in class i}},$$

$$\text{F-measure (i)} = \frac{2 \times \text{precision (i)} \times \text{recall (i)}}{\text{precision (i)} + \text{recall (i)}},$$ where $i = 1, 2$ with classes 1 and 2 being Web forum messages written by female and male authors respectively.

2.5 Experimental Study

To assess the effectiveness of the proposed research design, I conducted the experiment on a large and long-standing international Islamic women's political forum. In the following sections, detailed information about the testbed, hypotheses, experimental results, and discussion is provided.

2.5.1 Testbed

The experiment was conducted on a large, international Islamic women's political forum to evaluate the proposed framework of gender classification for Web forums. I chose it for three reasons: first, it is a large, long-standing (about 4 years), international political forum and thus can be used to study the international cyber political movement; second, it has self-reported gender information for each registered member, thus providing a gold-standard to evaluate the performance of the automatic gender classifiers; third, since it is a women's forum, more females participate, thus providing a larger number of messages written by female authors compared with other general, male-dominated Web forums. I believe the international, political, and female-oriented nature of this large active forum makes it an ideal testbed for this research study.

All the messages in the forum posted up to March, 2007 were collected and parsed. In total, 34,695 different messages in 4,352 unique threads were gathered. The numbers of messages written by females and males were quite balanced. The time span of the collected messages is from June, 2004 to March, 2007. Based on careful discussion with the political science collaborator, who has significant experience in studying women's political forums, I believe that this testbed is of high quality and has credible participant-specified gender information.

To test the performance of the classifiers, I randomly selected 100 authors, 50 females and 50 males. In total, there were 12,690 messages posted by those 100 authors. Table 2.1 shows the distribution of the numbers of messages written by females and males. In each number range, there were relatively balanced numbers of messages between females and males. On average, each female participant produced 142.26 messages, and each male participant wrote 111.54 messages.

Table 2.1 Testbed distribution between females and males

Number Range of Posted Messages	**Female**	**Male**
5-20 messages	16	18
21-50 messages	10	16
51-100 messages	6	4
101-200 messages	8	4
201-500 messages	5	6
501-1000 messages	4	1
1000+ messages	1	1
Total	**50**	**50**

2.5.2 Hypotheses

Drawing on the vast online social media classification literature, I posited that adding content-specific features to the baseline content-free features would improve the performance of gender classification for Web forums (targeting research question 2), and that conducting feature selection on a relatively large number of features would improve the performance of gender classification for Web forums (targeting research question 3). The specific hypotheses tested are as follows:

> H1: Using the combination of content-free features and unigrams can achieve higher performance than using only content-free features; i.e., F1+F2+F3+F4(unigram) > F1+F2+F3.
>
> H2: Using the combination of content-free features, unigrams, and bi-grams can achieve higher performance than using the combination of content-free features and unigrams; i.e., F1+F2+F3+F4(unigram)+F4(bi-gram) > F1+F2+F3+F4(unigram).
>
> H3: Using the feature set generated by conducting feature selection on the combination of content-free features and unigrams can achieve higher performance than using the combination of content-free features and unigrams without feature selection; i.e., selected F1+F2+F3+F4(unigram) > F1+F2+F3+F4(unigram).
>
> H4: Using the feature set generated by conducting feature selection on the combination of content-free features, unigrams, and bi-grams can achieve higher performance than using the combination of content-free features, unigrams, and bi-grams without feature selection; i.e., selected F1+F2+F3+F4(unigram)+F4(bi-gram) > F1+F2+F3+F4(unigram)+F4(bi-gram).
>
> H5: Using the feature set generated by conducting feature selection on the combination of content-free features, unigrams, and bi-grams can achieve higher performance than using the feature set created by conducting feature selection on the combination of content-free features and unigrams; i.e., selected F1+F2+F3+F4(unigram)+F4(bi-gram) > selected F1+F2+F3+F4(unigram).

2.5.3 Experimental Results

Feature set F1+F2+F3 contained 250 content-free features. For the content-specific features, 6,012 unigrams and 4,022 bi-grams were obtained. Therefore, there were 6,262 and 10,284 features in feature sets F1+F2+F3+F4(unigram) and F1+F2+F3+F4(unigram)+F4(bi-gram) respectively. After feature selection, the two selected feature sets consisted of 351 and 640 features respectively, each of which was much smaller than the corresponding one without feature selection. The feature selection was carried out by Weka's Information Gain attribute evaluator (Witten and Frank 2005).

The classification was carried out by using a linear kernel with the Sequential Minimal Optimization (SMO) algorithm (Platt 1999) included in the Weka Data Mining Package (Witten and Frank 2005). Evaluation was done via 10-fold cross validation. In each fold, 90% of the data was used as the training set and the remaining 10% as the testing set. Figure 2.2 shows the accuracy of gender classification on each feature set. Accuracy increased as more types of features were incorporated. Specifically, feature set F1+F2+F3+F4(unigram)+F4(bi-gram) outperformed feature set F1+F2+F3+F4(unigram), which in turn outperformed feature set F1+F2+F3. In addition, feature selection improved the classification accuracies significantly. Specifically, after conducting feature selection, the classification accuracies on feature sets F1+F2+F3+F4(unigram) and F1+F2+F3+F4(unigram)+F4(bi-gram) increased from 62% to 83% and 64% to 86% respectively. The highest classification accuracy was achieved on the selected feature set F1+F2+F3+F4(unigram)+F4(bi-gram).

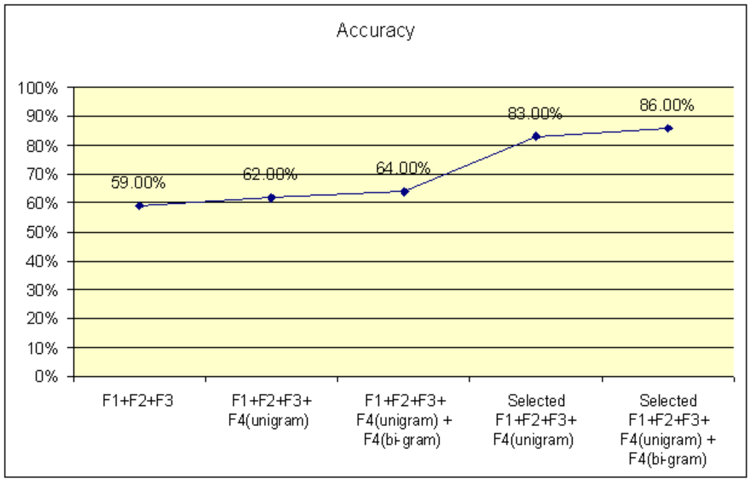

Figure 2.2 The accuracy of gender classification on each feature set

Table 2.2 shows the precision, recall and F-measure of gender classification on each feature set. All three types of measurement values increased in the same way as the accuracy (see Figure 2.2). The highest precision, recall, and F-measure for both classes (i.e., female and male) were achieved on the selected feature set F1+F2+F3+F4(unigram)+F4(bi-gram). Compared to Corney et al.'s study (2002) of gender classification on e-mails, the best F-measure (i.e., 86.45%) achieved in this study was higher than that (i.e., 71.1%) reported in their study. However, the highest accuracy (i.e., 91.5%) reported in Nowson and Oberlander's work (2006) on gender classification in the context of Web blog was higher than the highest accuracy (i.e., 86%) achieved in this study. This could be attributed to the shorter text in the Web forum messages compared with the Web blog corpus they used.

Table 2.2 Performance measures using different feature sets

Feature Set	Class	Precision	Recall	F-measure
F1+F2+F3	Female	57.10%	72.00%	63.69%
	Male	62.20%	46.00%	52.89%
	Average	59.70%	59.00%	59.35%
F1+F2+F3+F4(unigram)	Female	63.00%	58.00%	60.40%
	Male	61.10%	66.00%	63.46%
	Average	62.10%	62.00%	62.05%
F1+F2+F3+F4(unigram) +F4(bi-gram)	Female	62.50%	70.00%	66.04%
	Male	65.90%	58.00%	61.70%
	Average	64.20%	64.00%	64.10%
Selected F1+F2+F3+F4(unigram)	Female	90.20%	74.00%	81.30%
	Male	78.00%	92.00%	84.42%
	Average	84.10%	83.00%	83.55%
Selected F1+F2+F3+F4(unigram) +F4(bi-gram)	Female	92.90%	78.00%	84.80%
	Male	81.00%	94.00%	87.02%
	Average	86.90%	86.00%	86.45%

Pair-wise t-tests on accuracy and F-measure were performed to test H1 through H5. The tests were conducted by randomly shuffling the data before performing 10-fold cross validation; i.e., I reshuffled the data in a different way each time and then performed 10-fold cross validation. This process was repeated 30 times. As summarized in Table 2.3, the classification accuracy and F-measure were significantly higher ($p < 0.0001$ on both accuracy and F-measure) when using the combination of content-free features and unigrams than those using only content-free features. Thus, H1 is supported. Compared with the combination of content-free features and unigrams, the combination of content-free features, unigrams, and bi-grams showed significantly higher classification accuracy ($p = 0.0080$) and F-measure ($p = 0.0400$). Therefore, H2 is supported. After feature selection, the selected feature set F1+F2+F3+F4(unigram) outperformed feature set F1+F2+F3+F4(unigram) significantly ($p < 0.0001$ on both accuracy and F-measure), in support of H3; and the selected feature set F1+F2+F3+F4(unigram)+F4(bi-gram) outperformed feature set F1+F2+F3+F4(unigram)+F4(bi-gram) significantly ($p < 0.0001$ on both accuracy and F-measure), in support of H4. In addition, the classification accuracy and F-measure were significantly higher ($p < 0.0001$ on both accuracy and F-measure) when using the selected feature set F1+F2+F3+F4(unigram)+F4(bi-gram) than when using the selected feature set F1+F2+F3+F4(unigram). Thus, H5 is supported.

Table 2.3 Results of hypotheses testing on accuracy and F-measure for H1-H5

No.	***p* value on accuracy**	**Result**	***p* value on average F-measure**	**Result**
H1	<0.0001**	Supported	<0.0001**	Supported
H2	0.0080**	Supported	0.0400*	Supported
H3	<0.0001**	Supported	<0.0001**	Supported
H4	<0.0001**	Supported	<0.0001**	Supported
H5	<0.0001**	Supported	<0.0001**	Supported

Note. Significance levels * $\alpha = 0.05$ and ** $\alpha = 0.01$.

2.5.4 Discussion

This section provides detailed discussion on the performance of different feature sets.

Feature set F1+F2+F3 achieved 59.00% accuracy, 59.70% average precision, 59.00% average recall, and 59.35% average F-measure. However, compared with the results reported in previous online text classification studies (Abbasi and Chen 2005; Zheng et al. 2006), the performance results achieved in this study were worse. This could be attributed to several causes. First, as mentioned in previous studies, although they represent vocabulary richness, lexical features (F1) may not be very useful when the text length is short (Zheng et al. 2006). Since some Web forum messages in the dataset are quite short, the lexical features would not have been, after all, very effective. Second, compared with the relatively small number of words in a Web forum message, the 150 function words used in this study as part of the syntactic features (F2) may be more than necessary. In their study, de Vel et al. (2001) observed a decrease in performance when the number of function words increased from 122 to 320. Third, since this study used only five structural features (F3), other important ones that may have had an impact in previous studies might have been missed.

Feature set F1+F2+F3+F4(unigram) was significantly better than feature set F1+F2+F3 (Table 2.3). This result is consistent with the previous studies (Abbasi and Chen 2008; Zheng et al. 2006) that point out the good discriminating capability of content-specific features. Those studies have noticed that Web forum participants were interested in different topics, thus providing the content-specific features with relatively high discriminatory power.

Feature set F1+F2+F3+F4(unigram)+F4(bi-gram) was significantly better than feature set F1+F2+F3+F4(unigram) (Table 2.3). This result indicates that although unigrams are very important content-specific features, they may not be sufficient to represent the content. By incorporating bi-grams, more content information about the

Web forum messages can be captured. However, I also noticed that the increases of both precision and F-measure, although significant, were not as great as the increases from feature set F1+F2+F3+F4(unigram) to the baseline feature set F1+F2+F3. One possible reason could be that although the large number of bi-grams captured more content information, it introduced noise as well (Li et al. 2007; Meiri and Zahavi 2006).

Koppel & Schler (2003) have shown that conducting feature selection on n-grams can improve the text classification results. According to the t-test results of H3 and H4 (Table 2.3), feature selection improved the classification performance significantly. As shown in Figure 2.2 and Table 2.2, all four performance measures increased significantly after conducting feature selection on feature sets F1+F2+F3+F4(unigram) and F1+F2+F3+F4(unigram)+F4(bi-gram) respectively. In addition, the numbers of features in the selected feature sets were reduced appreciably, thus leading to higher efficiency.

The testing between the two selected feature sets (see H5) showed that the selected feature set F1+F2+F3+F4(unigram)+F4(bi-gram) outperformed the selected feature set F1+F2+F3+F4(unigram) significantly (Table 2.3). Similar to the comparison between feature sets F1+F2+F3+F4(unigram)+F4(bi-gram) and F1+F2+F3+F4(unigram), this result once again indicates that using both unigrams and bi-grams can better improve the gender classification performance for Web forums than using only unigrams as content-specific features.

2.5.5 Different Topics of Interests: Females and Males

In the experimental study, the highest classification accuracy of 86% was achieved on the selected feature set F1+F2+F3+F4(unigram)+F4(bi-gram). This indicates that gender differences do exist in Web forums and the features used for classification, especially the content-specific features, have a high discriminating capability of distinguishing the online gender differences between female and male posters, thus answering the research question 1.

By investigating the features in the selected feature set F1+F2+F3+F4(unigram)+F4(bi-gram), I observed that females talked more about family members, God, peace, marriage, and good will; on the other hand, males talked more about extremism, holy men, and belief.

Table 2.4 lists some examples of the unigrams and bi-grams preferred by females and males respectively from the selected feature set F1+F2+F3+F4(unigram)+F4(bi-gram). They are among the features with the highest Information Gain values, therefore showing high discriminatory power. Chi-square (χ^2) tests were conducted to examine the statistical significances of the differences in using those unigrams and bi-grams between females and males. A domain expert from an Islamic country provided the meanings of some of those unigrams and bi-grams.

Table 2.4 Examples of female and male preferred unigrams and bi-grams from the selected feature set F1+F2+F3+F4(unigram)+F4(bi-gram)

Female preferred unigrams and bi-grams			
Keyword	**χ^2 value**	***p* value**	**Meaning**
Sis	456.07	<0.0001**	Sisters in Islam
Sister	165.08	<0.0001**	
Mother	123.88	<0.0001**	
Husband	51.87	<0.0001**	
Flower	9.00	0.0030**	
Amen	166.64	<0.0001**	
Alhamdulillah	283.85	<0.0001**	Thank God
Inshaallaah	33.51	<0.0001**	In God’s will
Ahhah kheir	15.16	<0.0001**	God is good
Sexually defiled	5.25	0.0220*	
Male preferred unigrams and bi-grams			
Keyword	**χ^2 value**	***p* value**	**Meaning**
Salafi	377.17	<0.0001**	Extremist sect of Islam
Allah	290.30	<0.0001**	Allah God of Muslims
Army	66.12	<0.0001**	
Deviant	35.79	<0.0001**	
Ijtihaad	57.80	<0.0001**	Inferring or interpreting Islamic laws
E-mail	23.81	<0.0001**	
Great scholar	13.89	0.0002**	
Muslim intellectual	11.27	0.0008**	
Imam Nawawi	26.56	<0.0001**	Priest Nawawi
Original Arabic	3.52	0.0606	

Note. Significance levels * $\alpha = 0.05$ and ** $\alpha = 0.01$.

As summarized in Table 2.4, all the listed, female-preferred unigrams and bi-grams were statistically significant. Specifically, significant terms/words in female conversations included: sis (i.e., sisters in Islam), sister, mother, husband, flower, amen, alhamdulillah (i.e., thank God), inshaallaah (i.e., in God's will), ahhah kheir (i.e., God is good), and sexually defiled. Male-preferred unigrams and bi-grams were statistically significant, except for "original Arabic." Specifically, significant terms/words in male discussions included: Salafi (i.e., an extremist sect of Islam), Allah (i.e., Allah God of Muslims), army, deviant, ijtihaad (i.e., inferring or interpreting Islamic laws), e-mail, great scholar, Muslim intellectual, and imam Nawawi (i.e., Priest Nawawi). For the bi-gram "original Arabic," although men preferred to use it more frequently than women, the difference was not statistically significant ($p = 0.0606 > 0.05$). This may be because the total number of its appearances in the whole forum was small and therefore could not show statistical significance.

The results of the experimental study show the importance of content-specific features in gender classification for Web forums, and are consistent with previous gender classification studies for Web blogs (Nowson and Oberlander 2006; Schler et al. 2006).

As an important type of social media, political Web forums have become a major communication channel for people to discuss and debate political, cultural and social issues. More and more women are using this medium to share their political opinions and knowledge. Along with this trend, researchers have developed an increased interest in studying online gender differences. By analyzing writing styles and topics of interest, the experimental results indicate that female and male participants in political Web forums do have significantly different topics of interest.

2.6 Conclusions and Future Directions

2.6.1 Research Contributions

With the rapid development and the increasing importance of the Internet in people's daily lives and work, understanding online gender differences and why they occur is becoming more and more important for Internet service providers, system developers, information analysts, and end users.

Nowadays, more and more women are participating in cyberspace. However, this does not diminish online gender differences. In contrast, discrepancies in motivation and interest in Internet use between females and males are becoming the focus of online gender difference research. This study used feature-based online social media text classification techniques to investigate the online gender differences between female and male participants in Web forums, by examining their writing styles and topics of interest. The proposed feature-based gender classification framework can be applied to other different Web forums.

In the framework, I examined different types of features that have been widely used in previous online text classification studies, including: lexical features, syntactic features, structural features, and content-specific features. For content-specific features, unigrams and bi-grams automatically extracted from the whole forum were used instead of manually selected. Five different feature sets were built by adding content-specific features to the basic content-free features and conducting feature selection. In the experimental study on a large Islamic women's political forum, the feature sets combining both content-specific and content-free features performed significantly better than the ones consisting of only content-free features. In addition, feature selection on large feature sets improved the classification performance significantly. The results also indicated the existence of online gender differences in Web forums. Through further investigation on this selected feature set, different topics of interest between females and males were identified.

This research has made several contributions. First, a systematic framework of gender classification was proposed to analyze online gender differences in social media, an area which has received little investigative attention. The framework can be applied to study the gender differences in many other domains (e.g., sociology, business, and marketing). It provides an informative point of departure for continued research. Second, the empirical study demonstrated the effectiveness of the proposed framework, thus confirming the prevalence of online gender differences in Web forums. Third, this study also makes a research contribution by examining different feature sets and identifying the one with the best classification performance. The comparison of different feature sets also indicates the importance of incorporating content-specific features and conducting feature selection in automatic gender classification for online social media.

2.6.2 Future Research Directions

This study also has some limitations that can be explored further. First, this study used only unigrams and bi-grams as content-specific features. Because of their computational complexity, I did not include n-grams with n greater than 2. However, those n-grams could capture more content information than unigrams and bi-grams, thus potentially leading to higher classification performance. On the other hand, if n is too big, the n-grams may introduce additional noise and computational overhead. More systematic investigation may be needed. Second, to generate the selected feature sets, I adopted Information Gain which is one of the most widely used feature selection methods. However, there are other advanced feature selection methods such as the wrapper model, the filter model, and the Markov blanket. Future research could explore and compare their performance in gender classification in the Web forum context. Third, the proposed gender classification framework was tested on only one English language forum. I believe the framework can be applied to Web forums in different languages, but feature representation and extraction research would need to be conducted to better develop a scalable, multilingual online gender classification model. Lastly, I plan to explore the gender differences in other important social media domains, such as marketing, e-commerce, health care, and education. These are all areas in which women may exhibit unique characteristics and exercise significant influence. Additional social, cultural, and psychological models would also then need to be considered in future research.

CHAPTER 3. EXAMINING THE EMOTIONAL DIFFERENCES BETWEEN WOMEN AND MEN IN WEB FORUM COMMUNICATION

3.1 Introduction

As discussed in Chapter 2, with the increasing use of online social media, questions concerning gender differences in the new media have been raised. In Chapter 2, a feature-based text classification framework was proposed to examine online gender differences between Web forum posters by analyzing writing styles and topics of interest. In this chapter, further effort is made to examine the emotional differences between the two genders in social media by developing a research framework using sentiment analysis techniques.

Through his influential 1992 bestselling book, *Men Are From Mars, Women Are From Venus,* John Gray (1992) asserted that women and men are quite different in their communication styles and emotional needs. Although later criticisms (such as that in *The Myth of Mars and Venus: Do Men and Women Really Speak Different Languages?* by Deborah Cameron (2007)) note that the book is almost exclusively about differences and there is as much similarity and variation within each gender as between women and men, the critics also agree that differences in emotional communication between the two genders do exist.

Previous research on gender differences in emotion has shown that women are more emotional than men (Fischer and Manstead 2004; Thelwall et al. 2010). In addition, many studies have found that women tend to express more intensely positive emotions such as happiness, love, and life satisfaction than men (Fujita et al. 1991; Newman et al. 2008; Wood et al. 1989). In terms of negative emotions, although some inconsistencies exist, women have reported higher levels of negative affect and depression (Gove 1978; Grossman and Wood 1993; Nolen-Hoeksema 1987), as well as greater fear and sadness (Scherer et al. 1986), whereas men are reported to express more anger (Lucas and Gohm 2000). Theories of stereotyping (Brebner 2003; Brody 1997) and social roles (Eagly 1987; Eagly and Wood 1991) have shed light on the potential explanations of such emotional differences between the two genders.

However, many of the studies on gender differences in emotion focus on face-to-face settings. Few have specifically examined such differences in the online world. With the advent of Web 2.0 (O'Reilly 2005), a large number of people began to participate and became active in text-based CMC to exchange information and express their opinions. The amount of user-generated content in these media has been experiencing exponential growth. Early research assumed that text-based CMC could not transmit socio-emotional content (Sproull and Kiesler 1986). However, later studies using social information processing theory (Walther 1992) argued that text-based CMC could support socio-emotional and relational communication. Therefore, the users of the new media are likely to adapt existing face-to-face communication cues in the physical world to the online world (Guiller and Durndell 2007; Jaffe et al. 1999).

To examine this problem, a generic framework was proposed in this chapter to compare the intensity of emotions between the two genders, specifically on subjectivity vs. objectivity, and positivity vs. negativity, using sentiment analysis techniques. An experiment was conducted on a large and long-standing international women's political forum suggested by an expert researcher in women's studies.

The remainder of this chapter is organized as follows. Section 3.2 reviews gender differences in emotion and the related theoretical explanations, the emotional differences in text-based CMC, and sentiment analysis techniques that can be used to analyze emotional differences. Section 3.3 summarizes the literature review and highlights the motivation for this study. Section 3.4 provides detailed research design in terms of data acquisition and emotion detection. Section 3.5 describes the experimental study and highlights key results, followed by a discussion of the study's contributions and some future directions in Section 3.6.

3.2 Theoretical Background and Enabling Techniques

3.2.1 Gender Differences in Emotion and Related Theoretical Explanations

As suggested by the previous reviews on gender differences in emotion, women are more emotional than men and they are more likely to express positive emotions than men (Fischer and Manstead 2004). Although some inconsistencies exist for negative emotions, in general, women are reported to express more sadness, fear, shame, and guilt, and men are reported to have more anger and other hostile emotions (Fischer and Manstead 2004).

Previous research on gender differences in happiness and satisfaction found that women tend to express emotions more than men (Brody and Hall 1993; Wood et al.

1989), are more affectionate (Briton and Hall 1995), and experience more intense joy and sadness (Fujita et al. 1991). Previous research has also found that women tend to express more warmth and concern for others, as well as higher levels of happiness and life satisfaction (Grossman and Wood 1993). In addition, women are also reported to have higher levels of negative emotions such as depression, fear, and sadness (Grossman and Wood 1993; Scherer et al. 1986). A cross-cultural analysis by Lucas and Gohm (2000) showed that women express both negative emotions (especially fear and sadness) and positive emotions more frequently than men.

Many researchers have suggested that any differences between women and men are the result of stereotyping and social roles (Brebner 2003; Brody 1997; Eagly 1987; Eagly and Wood 1991). Gender differences in emotion have generally been influenced by social and cultural context during the gender-stereotypic socialization process (Fischer and Manstead 2004).

Brody (1997) argues that stereotypes provide models of appropriate behavior that people may adopt. Emotional differences between the two genders may be attributed to the differences in stereotypes of women and men (Brebner 2003). One of the most consistent gender stereotypes found by researchers is that women are more emotional than men (Thelwall et al. 2010). Some studies have found that although women and men experience most emotions to a similar degree, women tend to express sadness, fear, and love more frequently while men tend to express anger more frequently (Fabes and Martin 1991; Plant et al. 2000; Stoppard and Gunn-Gruchy 1993). Other studies have also found that women tend to smile more and express more warmth than men (Briton and Hall

1995; Brody 1997). These stereotypes can heavily influence gender differences in emotion. People who do not conform to stereotypic behavior may be punished through social rejection, whereas people who do conform may be encouraged or rewarded for such behaviors in the form of social approval (Brody 1997; Fiske and Stevens 1993).

Emotions can be considered as part of the role socialization process for women's and men's occupations (Fischer and Manstead 2004; Shields et al. 2007). Wood and Eagly (2002) study gender differences in social behavior and emphasize the importance of social roles in explaining such differences. Their analysis indicates that gender differences may be because of the interaction between physical attributes and social arrangements in society. Traditionally, women are more likely than men to take caretaker roles. For example, they are taking the roles of wife and mother at home and teacher and nurse at work. On the contrary, men are more likely to work in employment settings that involve providing material resources (Fischer and Manstead 2004; Grossman and Wood 1993). These role differences can then lead to gender differences in social behavior since these roles can generally construct the beliefs, social stereotypes, and expectations of behavioral differences between women and men (Grossman and Wood 1993). By filling their different role expectations, different emotional expressions are needed for the two genders (Fischer and Manstead 2004). Because of their caretaker roles, women are becoming sensitive to others' needs and emotions; however, men's roles do not have such requirements (Grossman and Wood 1993). Therefore, during role construction, women learn to be more emotional, and men to be less emotional (Brebner 2003). Previous studies have hypothesized and found that women are characterized by greater sensitivity to nonverbal signals, expressions of vulnerability, weakness, and helplessness than are men; however, men need to deny such motional experiences because of the widely accepted male identity such as achievement (Brody and Hall 2008; Brody 1985; Miller 1976; Thelwall et al. 2010).

3.2.2 Gender Differences in Emotion in Text-Based CMC

The increased use of networked computers in contemporary society has changed the ways in which people communicate with others. With the advent of Web 2.0 (O'Reilly 2005), computer-mediated communication (CMC) enabled all kinds of communication to take place online. An individual can conduct more interactive communication and act positively using this new medium instead of only passively acquiring information. The two-way communication between end-users and the Web-based communities enhances the information and opinion sharing among Internet users. Examples of CMC include e-mails, Web forums, blogs, wikis, social-networking sites, chat rooms, instant messages, etc. There are two types of CMC, termed synchronous CMC and asynchronous CMC (Guiller and Durndell 2007). Synchronous CMC, such as chat rooms, takes place in real time. In contrast, asynchronous CMC refers to communication that takes place over computers in delayed time (e.g., emails, Web forums, blogs, wikis). Hence, asynchronous CMC allows people to communicate from different places and at different times. The major advantage is that participants can take their time to read and respond to messages (Guiller and Durndell 2007; Harasim 1990).

Previous psychological research has argued that the Internet could reduce both self-awareness and the transmission of interpersonal information, because less social context cues can be expressed in the text-based, asynchronous CMC compared with the face-to-face communication in the physical world (Guiller and Durndell 2007). However, social information processing theory (Walther 1992) suggests that text-based CMC could support socio-emotional communication since users are affected by the same level of awareness of interaction with others as in face-to-face communication (Jaffe et al. 1999). Therefore, it is suggested that CMC users will adapt existing communication cues in the physical world to the online world, even though there could be restrictions of language use and display in the online world (Abbasi et al. 2008a; Guiller and Durndell 2007).

Moreover, gender may also influence the socio-emotional communication in CMC (Guiller and Durndell 2007). Men are often treated as independent problem solvers; therefore they tend to emphasize competition more in their communications. However, women tend to value collaboration and affiliation more than men, and emphasize cooperation in their communications (Coates 1993; Tannen 1991). Therefore, women may be more likely than men to express socio-emotional opinions and personal feelings in text-based CMC (Guiller and Durndell 2007).

3.2.3 Enabling Techniques for Sentiment Analysis

Sentiment analysis is the automatic detection of opinions from free text. The computer-based detection and analysis of emotion, particularly in text, has been a growing interest among researchers in recent years (Pang and Lee 2008). Previous sentiment analysis studies include determining whether a text is objective or subjective (Pang et al. 2002; Turney 2002), or whether a subjective text contains positive or negative sentiments (Wiebe et al. 2001; Wiebe et al. 2004). Most previous sentiment analysis research focuses on classifying an opinioned piece of text into one of two opposing sentiment polarities (Pang and Lee 2008). In addition, other studies have attempted to determine if a piece of text expresses happiness, sadness, anger, horror, etc. (Grefenstette et al. 2004; Mishne 2005; Subasic and Huettner 2001).

Two approaches in sentiment analysis have been widely used: the machine learning approach and the semantic orientation approach (see (Liu 2007; Pang and Lee 2008; Zhou and Chaovalit 2008) for more comprehensive reviews). The machine learning approach treats sentiment analysis as a topic-based text classification problem. Any text classification algorithm can be employed, e.g., Naïve Bayes (Gamon et al. 2005; Pang et al. 2002), Support Vector Machines (SVM) (Dave et al. 2003; Pang et al. 2002), Maximum Entropy (Chaovalit and Zhou 2005; Pang et al. 2002) etc. For example, Pang et al. (2002) compared the performances of Naïve Bayes, SVM, and Maximum Entropy using movie reviews and found that Naïve Bayes and SVM performed well in determining if a piece of a review was negative or positive. Typically, a classifier is built on the training data, and is then evaluated on the testing data to fine tune its performance.

Features are important when using this approach. A piece of text is usually converted into a feature vector that can represent the most salient and important information expressed in the original text. Different types of features have been used in sentiment classification studies, such as bag-of-words, n-grams (Abbasi et al. 2008b;

Gamon 2004; Pang et al. 2002), Part-of-Speech (POS) tags (Abbasi et al. 2008b; Gamon 2004; Pang et al. 2002), and word position (Hovy 2006; Pang et al. 2002). As summarized by Pang and Lee (2008), whether higher-order n-grams are useful features appears to be a matter of some debate. For example, Pang et al. (2002) reported that unigrams outperformed bigrams when classifying movie reviews by sentiment polarity, but Dave et al. (2003) found that in some settings, bigrams and trigrams yielded better product-review polarity classification. POS is commonly exploited in sentiment analysis because it can be considered to be a crude form of word sense disambiguation (Wilks and Stevenson 1998). Adjectives have been widely accepted as good POS features (Pang and Lee 2008). Other words, such as nouns, verbs, and adverbs have also been used (Benamara et al. 2007; Wiebe et al. 2004). The position of a word within a textual unit (e.g., in the middle vs. near the end of a document) can potentially affect how much that word affects the overall sentiment or subjectivity status of the enclosing textual unit (Pang and Lee 2008). Thus, position information is sometimes encoded into the feature vectors that are employed (Hovy 2006; Pang et al. 2002). In addition, some studies have also utilized complex features of phrase patterns, which make use of POS and n-gram patterns (Fei et al. 2004). For example, Lei et al. (2004) noted that phrase patterns such as “n+aj” (noun followed by positive adjective) typically represented positive sentiment orientation while “n+dj” (noun followed by negative adjective) often expressed negative sentiment.

Overall, the machine learning approach tends to be more accurate than the semantic orientation approach since a machine learning model is always tuned to the training data set; however, this also becomes a disadvantage of the machine learning approach as it is domain dependent with less generalizability (Liu 2007; Turney and Littman 2003; Zhou and Chaovalit 2008). If applied elsewhere, training on the new data sets is needed.

In contrast to the machine learning approach, the semantic orientation approach is domain independent with better generalizability (Liu 2007; Turney and Littman 2003; Zhou and Chaovalit 2008). This approach performs classification based on positive and negative sentiment words and phrases contained in each evaluation text and no prior training is required in order to mine the data (Pang and Lee 2008). This approach often relies on external knowledge resources beyond raw data and thus is knowledge-rich (Zhou and Chaovalit 2008). It generally proceeds in three steps: (1) extracting words or phrases that express semantic orientations; (2) determining the polarities of the extracted words or phrases; and (3) computing the polarity of the text by aggregating the polarities of individual words or phrases in the text (Zhou and Chaovalit 2008).

Most semantic orientation-based sentiment analysis research has followed the procedure of first creating a sentiment lexicon and then calculating the positivity (or subjectivity) scores of a given textual document by mapping the positive and negative (or subjective) words and phrases with the corresponding entries in the lexicon (Pang and Lee 2008). For example, Turney (2002) calculated a phrase's semantic orientation to be the mutual information between the phrase and the word "excellent" (as the positive polarity), minus the mutual information between the phrase and the word "poor" (as the negative polarity). The overall polarity of an entire text was predicted as the average semantic orientation of all the phrases that contained adjectives or adverbs. Hu & Liu (2004) used a bootstrapping technique to generate a set of opinion words with semantic orientations from a group of manually created seed adjectives, by searching for synonyms and antonyms in WordNet. The orientation of a sentence was determined by the dominant orientation of the opinion words in the sentence. That is, if positive (negative) opinion prevailed, the sentence was regarded as a positive (negative) one. Building upon WordNet, SentiWordNet (Esuli and Sebastiani 2006) is a lexical resource for sentiment analysis which has more sentiment related features than WordNet. It assigns to each synset of WordNet three sentiment scores regarding positivity, negativity, and objectivity respectively. SentiWordNet has been used as the lexicon in recent sentiment classification studies (Dang et al. 2010; Denecke 2008; Devitt and Ahmad 2007; Fahrni and Klenner 2008).

3.3 Research Motivation

In traditional face-to-face communication, it is generally concluded that women tend to be more emotional than men. In addition, many studies have indicated that women are more likely to express positive emotions (such as happiness and satisfaction) than men. Although some inconsistencies exist, most studies have also found that women are more likely to report negative emotions (such as fear and sadness, but not anger) as well. Stereotyping and social roles theories have proposed potential explanations for the emotional differences between the two genders. In online communication settings, some studies on CMC, especially text-based, asynchronous CMC, have hypothesized that users will adapt their existing face-to-face communication cues to the new medium and therefore arguing that emotions expressed in text-based CMC will not reduce in comparison to the communication in the physical world. However, few studies have explicitly examined the emotional differences between women and men in text-based CMC; the only study I have found was done by Thelwall et al. (2010) that compared the emotional differences between women and men in MySpace (i.e., a social network site) using random sampling and a manual analysis method.

In this study, a generic framework is developed to automatically examine the emotional differences between women and men. In addition, to obtain a comprehensive, unbiased analysis result, longitudinal textual data from a whole text-based CMC site has been used in the framework instead of random sampling, which uses only a portion of the data. Moreover, since no studies have examined the emotional differences between the two genders in the Web forum context, an empirical experiment was conducted using the proposed automatic framework on a large and long-standing international women's political forum which was suggested by a social scientist in women's studies. The next section describes the research design in detail.

3.4 Research Design

To examine gender differences in emotion in text-based CMC, a research framework based on sentiment analysis techniques was developed. As shown in Figure 3.1, the framework includes three components: Data Acquisition, Emotion Detection, and Gender Comparison. The following subsections provide detailed description about the first two components. For the last component, the pairwise t-tests were conducted to examine the statistical significance of the emotional differences between women and men. The empirical testing results are presented in Section 3.5.

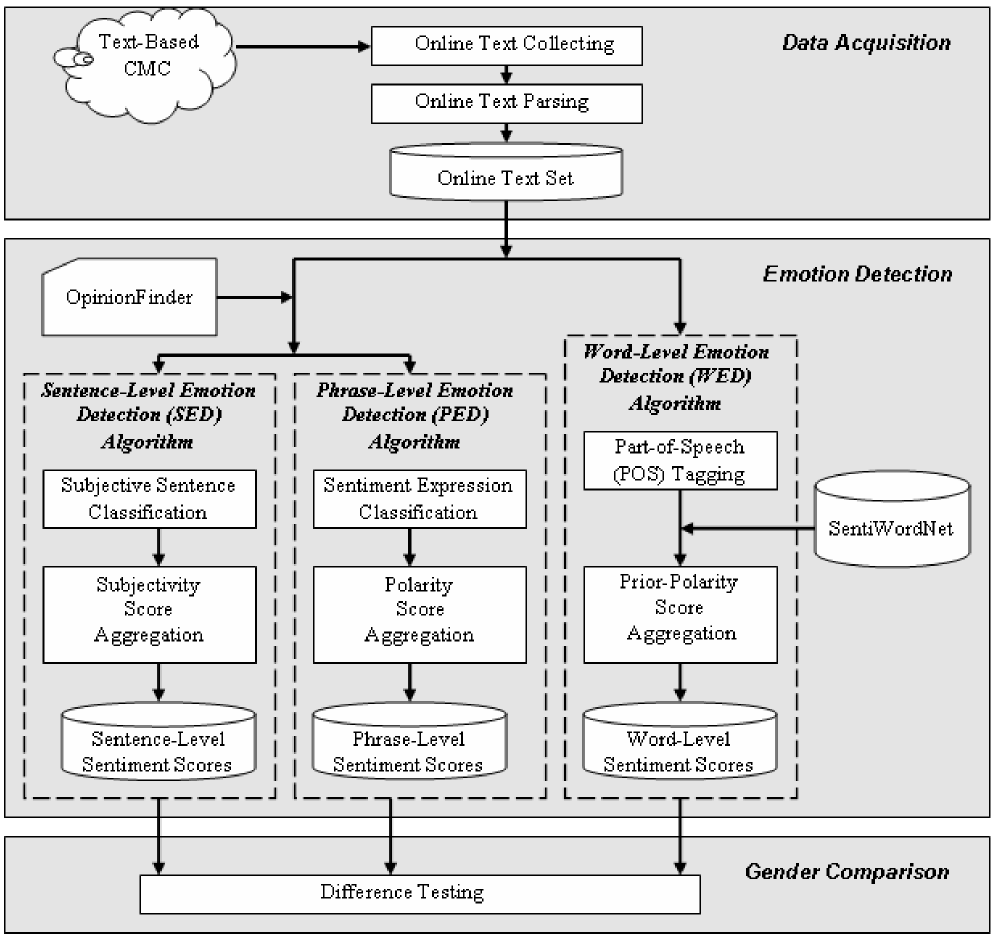

Figure 3.1 Research design of examining gender differences in emotion in text-based CMC

3.4.1 Data Acquisition

The data acquisition component consists of two steps: online text collecting and online text parsing. First, spidering programs were developed to collect the text from certain text-based CMC sites as HTML pages. After that, parsers were built to parse out the text body from the raw HTML pages and store the parsed data in a relational database.

3.4.2 Emotion Detection

In the emotion detection component, automatic sentiment analysis techniques are used to calculate the sentiment scores of the textual data generated by women and men respectively.

In text-based CMC, the contents are often longitudinal. The messages generated by different users can be viewed as a sequence of events or conversations. Therefore, to compare the emotional differences between women and men over time, these sequential events were aggregated into larger time units. According to the characteristics of a given text-based CMC data collection, the width of the time units can be days, weeks, months, quarters, years, etc.

To make the analysis more systematic, I adopted both machine learning and semantic orientation approaches using OpinionFinder and SentiWordNet, respectively. OpinionFinder is a well-known sentiment classification tool developed by Wiebe’s group (Riloff and Wiebe 2003; Wiebe and Riloff 2005; Wilson et al. 2005a). In this study, I used two major components of the tool: (1) the subjective sentence classification component, and (2) the sentiment expression classification component (Wilson et al. 2005a) to conduct sentence-level and phrase-level analyses respectively. As a well-known sentiment lexicon, SentiWordNet has been widely used in recent semantic orientation based sentiment classification studies (Dang et al. 2010; Denecke 2008; Devitt and Ahmad 2007; Fahrni and Klenner 2008). In this study, it was used do conduct the word-level analysis.

Sentence-Level Emotion Detection (SED) Algorithm. To perform sentence-level emotion detection, the Sentence-Level Emotion Detection (SED) algorithm was developed. First, the two classifiers in the subjective sentence classification component of OpinionFinder was used to classify each sentence in the whole data collection to be subjective, objective, or unknown. The classifiers are based on Naïve Bayes algorithm using a variety of lexical and contextual features (Riloff and Wiebe 2003; Wiebe and Riloff 2005) to distinguish between subjective and objective sentences. A set of rule-based features were also incorporated to increase the generalizability of the classifiers (Wilson et al. 2005a). The first classifier (C1) tags each sentence as either subjective or objective. This classifier uses the strategy that yields out highest overall accuracy, meaning the percentage of the right answers it achieves (according to manual annotations). Evaluated on 9,732 sentences from the MPQA opinion corpus (4,352 objective and 5,380 subjective sentences), this classifier had an accuracy of 74%, subjective precision of 78.4% (i.e., 78.4% of the sentences that the system classifies as subjective are indeed subjective, according to the manual annotations), subjective recall of 73.2% (i.e., of all the subjective sentences, 73.2% are automatically classified as subjective, rather than objective, or unknown), and subjective F-measure of 75.7%. The second classifier (C2) optimizes precision at the expense of recall. That is, it classifies a sentence as subjective or objective only if it can do so with confidence. Otherwise, it labels the sentence as “unknown.” Evaluated on the same 9,732 sentences from the MPQA opinion corpus, this strategy yielded 91.7% in subjective precision, 30.9% in subjective recall, 83.0% in objective precision, and 32.8% in recall.

Second, the aggregated subjectivity scores were calculated using the procedure shown in Figure 3.2. Within each time unit, only the sentences classified as subjective (objective) by both classifiers are kept. All the others are treated as “unknown.” Since the total numbers of sentences produced by women and men within a given time unit could be significantly different, the score is normalized by taking into account such total numbers. Specifically, for each time unit, the sentence-level subjective (objective) score for women (men) was calculated as the number of sentences classified as subjective (objective) by both C1 and C2 divided by the total number of sentences produced by women (men) within that time unit.

For each $t_i \in$ T (T is a set of text produced within a certain time unit)
For each $s_j \in t_i$ (s_j is a given sentence)
If $produced(t_i)$ = Women
N_{women} ++
Else
N_{men}++

If $C_1(s_j) \cap C_2(s_j)$ = Subjective AND $produced(t_i)$ = Women
$N_{women\text{-}subjective}$ ++
Else If $C_1(s_j) \cap C_2(s_j)$ = Objective AND $produced(t_i)$ = Women
$N_{women\text{-}objective}$ ++
Else If $C_1(s_j) \cap C_2(s_j)$ = Subjective AND $produced(t_i)$ = Men
$N_{men\text{-}subjective}$ ++
Else If $C_1(s_j) \cap C_2(s_j)$ = Objective AND $produced(t_i)$ = Men
$N_{men\text{-}objective}$ ++

$Score_{women\text{-}subjective} \leftarrow N_{women\text{-}subjective} / N_{women}$
$Score_{women\text{-}objective} \leftarrow N_{women\text{-}objective} / N_{women}$
$Score_{men\text{-}subjective} \leftarrow N_{men\text{-}subjective} / N_{men}$
$Score_{men\text{-}objective} \leftarrow N_{men\text{-}objective} / N_{men}$

Figure 3.2 The procedure of subjectivity score aggregation (Sentence-level Emotion Detection Algorithm)

Phrase-Level Emotion Detection (PED) Algorithm. To perform phrase-level emotion detection, the Phrase-Level Emotion Detection (PED) algorithm was developed. First, for the whole data collection, the sentiment expression classification component of OpinionFinder was used to identify words contained in phrases that express positive or negative sentiments following a two-step analysis (Wilson et al. 2005a). The first step classifies each phrase as neutral or polar using BoosTexter machine learning algorithm (Schapire and Singer 2000). The second step takes all phrases marked as polar in the first step to disambiguate their contextual polarity (Wilson et al. 2005b). Features used in the neutral-polar classifier include word features, modification features, sentence features, structure features, and document features. Features used in the polarity classifier include word features and polarity features. Evaluated on the MPQA opinion corpus, this component achieved an overall accuracy of 73.9% which is significantly better than the baseline majority rating method (Wilson et al. 2005b).

After that, the aggregated polarity scores were calculated using the procedure shown in Figure 3.3. Similar to the sentence-level subjectivity score aggregation, for each time unit, a phrase-level polarity score was calculated as the number of phrases that are classified as positive (negative) divided by the total number of sentences produced by women (men) in that time unit. The total number of sentences is used instead of the total number of phrases because only the phrases with polarities are detected. In addition, it is difficult to decide the scope of a phrase in general, since several short phrases may also be treated as a long phrase. Therefore, the total number of phrases cannot be exactly calculated. The total number of sentences is then used as an approximation to normalize the scores.

For each $t_i \in$ T (T is a set of text produced within a certain time unit)
For each $s_i \in t_i$ (s_j is a given sentence)
If $produced(t_i)$ = Women
N_{women} ++
Else
N_{men}++

For each $p_i \in t_i$ (p_j is a given phrase)
If $positive(p_j)$ = TRUE AND $produced(t_i)$ = Women
$N_{women\text{-}positive}$ ++
Else If $negative(p_j)$ = TRUE AND $produced(t_i)$ = Women
$N_{women\text{-}negative}$ ++
Else If $positive(p_j)$ = TRUE AND $produced(t_i)$ = Men
$N_{men\text{-}positive}$ ++
Else If $negative(p_j)$ = TRUE AND $produced(t_i)$ = Men
$N_{men\text{-}negative}$ ++

$Score_{women\text{-}positive} \leftarrow N_{women\text{-}positive} / N_{women}$
$Score_{women\text{-}negative} \leftarrow N_{women\text{-}negative} / N_{women}$
$Score_{men\text{-}positive} \leftarrow N_{men\text{-}positive} / N_{men}$
$Score_{men\text{-}negative} \leftarrow N_{men\text{-}negative} / N_{men}$

Figure 3.3 The procedure of polarity score aggregation (Phrase-level Emotion Detection Algorithm)

Word-Level Emotion Detection (WED) Algorithm. To perform word-level emotion detection, the Word-Level Emotion Detection (WED) algorithm was developed. As a newly developed sentiment-based lexicon, SentiWordNet has been used in recent sentiment classification studies. It assigns to each synset in WordNet three sentiment scores regarding positivity, negativity, and objectivity respectively. To use it, POS tagging on the whole data collection was first conducted. Once the POS tag for each word was decided, the sentiment score of a word was then calculated by looking up

SentiWordNet. According to previous literature (Benamara et al. 2007; Hatzivassiloglou and Wiebe. 2000; Hu and Liu 2004), adjectives, adverbs, verbs, and nouns were included in the analysis. Since each word in SentiWordNet has multiple POS senses, each of which is also related to multiple synsets, the average positive and average negative scores for its adjective, adverb, verb, and noun senses were calculated separately using the prior-polarity formula adopted from previous literature (Dang et al. 2010; Denecke 2008; Fahrni and Klenner 2008).

$$Score(word = pos)_i = \frac{\sum_{k \in SentiWordNet(word = pos \,\&\, polarity = i)} SentiWordNet\_Score(k)_i}{|synsets(word = pos)|}$$

Here, $pos \in \{\text{adjective}, \text{adverb}, \text{verb}, \text{noun}$, $i \in \{\text{positive}, \text{negative}\}$, and k denotes the synsets of a given word in a particular POS sense. To determine the polarity of each word in a given POS sense, the average positive score and the average negative score of the word were compared. It would be treated as positive if its average positive score was greater than its average negative score, and vice verse. For each time unit, a word-level polarity score regarding each POS sense (including adjective, adverb, verb, and noun) was calculated as the number of words that are identified as positive (negative) divided by the total number of words produced by women (men) in that time unit. Figure 3.4 summarizes the prior-polarity score aggregation procedure.

For each $t_i \in$ T (T is a set of text produced within a certain time unit)
For each $w_j \in t_i$ (w_j is a given word)
While $pos \in$ {adjective, adverb, verb, noun}
If $produced(t_i)$ = Women
$N(pos)_{women}$ ++
Else
$N(pos)_{men}$++

If $Score(word{=}pos)_{positive} > Score(word{=}pos)_{negative}$
AND $produced(t_i)$ = Women
$N(pos)_{women\text{-}positive}$ ++
Else If $Score(word{=}pos)_{positive} < Score(word{=}pos)_{negative}$
AND $produced(t_i)$ = Women
$N(pos)_{women\text{-}negative}$ ++
Else If $Score(word{=}pos)_{positive} > Score(word{=}pos)_{negative}$
AND $produced(t_i)$ = Men
$N(pos)_{men\text{-}positive}$ ++
Else If $Score(word{=}pos)_{positive} < Score(word{=}pos)_{negative}$
AND $produced(t_i)$ = Men
$N(pos)_{men\text{-}negative}$ ++

$Score(pos)_{women\text{-}positive} \leftarrow N(pos)_{women\text{-}positive} \,/\, N(pos)_{women}$
$Score(pos)_{women\text{-}negative} \leftarrow N(pos)_{women\text{-}negative} \,/\, N(pos)_{women}$
$Score(pos)_{men\text{-}positive} \leftarrow N(pos)_{men\text{-}positive} \,/\, N(pos)_{men}$
$Score(pos)_{men\text{-}negative} \leftarrow N(pos)_{men\text{-}negative} \,/\, N(pos)_{men}$

Figure 3.4 The procedure of prior-polarity score aggregation (Word-level Emotion Detection Algorithm)

3.5 Experimental Study

3.5.1 Data Set

To demonstrate the proposed framework for analyzing gender differences in emotion in text-based CMC, an empirical study was conducted in the Web forum context. As a major type of text-based CMC, Web forums provide the general public a platform to exchange information and express opinions with each other. Any participant can communicate with others in Web forums, no matter if they are acquainted or not.

A large and long-standing international women's political forum suggested by a women's studies expert was used as the testbed. I believe it is an ideal testbed for this study for the following reasons. First, the forum has self-reported gender information for each registered member, which can be used to distinguish the messages generated by women versus by men. Second, since it is a women's forum, the percentage of women participants are higher than that for other general, male-dominated Web forums, thus providing relatively balanced numbers of messages written by women and men. Third, there are a number of discussions and debates related to family issues, human rights, and political issues, thus making the content ideal for examining emotional differences.

In total, 34,695 messages in 4,352 different threads were collected. Among them, a few messages did not have either gender information or posted date information. After filtering out such messages, the remaining testbed contained 15,479 and 16,791 messages written by women and men respectively. The two numbers are quite balanced. The time span of all the collected messages is from June 9, 2004 to March 13, 2007, containing 34 months. For the analysis, the width of the time units was chosen to be one month. All messages in the same month are aggregated to calculate the sentiment scores using the SED, PED, and WED algorithms described in the previous section. Table 3.1 lists the numbers of messages broken down by gender and by month. For each month, the numbers of messages generated by women and men are generally balanced, except for October and November 2006, during each of which there were almost twice as many messages posted by men than by women. On average, women produced 455.26 messages per month with a similar number, 493.85, for men.

Table 3.1 Data set distribution by gender and by month

No.	**Month**	**Number of Messages**		**No.**	**Month**	**Number of Messages**	
		Generated by Women	**Generated by Men**			**Generated by Women**	**Generated by Men**
1	Jun, 2004	231	378	19	Dec, 2005	393	434
2	Jul, 2004	259	275	20	Jan, 2006	518	499
3	Aug, 2004	241	176	21	Feb, 2006	477	434
4	Sep, 2004	266	280	22	Mar, 2006	620	414
5	Oct, 2004	267	272	23	Apr, 2006	398	542
6	Nov, 2004	566	841	24	May, 2006	584	635
7	Dec, 2004	214	144	25	Jun, 2006	560	467
8	Jan, 2005	217	251	26	Jul, 2006	714	744
9	Feb, 2005	415	370	27	Aug, 2006	484	482
10	Mar, 2005	492	525	28	Sep, 2006	578	410
11	Apr, 2005	494	427	29	Oct, 2006	610	1,087
12	May, 2005	347	358	30	Nov, 2006	764	1,480
13	Jun, 2005	452	427	31	Dec, 2006	542	588
14	Jul 2005	318	385	32	Jan, 2007	636	710
15	Aug, 2005	377	433	33	Feb, 2007	661	692
16	Sep, 2005	515	497	34	Mar, 2007	406	309
17	Oct, 2005	305	375		**Total**	15,479	16,791
18	Nov, 2005	558	450		**Average**	455.26	493.85

3.5.2 Results and Discussion

The SED and PED algorithms were used to calculate the sentence- and phrase-level sentiment scores respectively. For each month, the sentence-level subjectivity scores and phrase-level polarity scores were aggregated using the procedures shown in Figures 3.2 and 3.3.

The WED algorithm was used to calculate the word-level sentiment scores. First, the Stanford POS tagger (http://nlp.stanford.edu/software/tagger.shtml) was used to tag the whole data collection. Only words that were tagged as adjective, adverb, verb, or noun were kept for further analysis. Then the prior-polarity scores for each word were calculated by looking up SentiWordNet and using the abovementioned prior-polarity formula. Finally, for each POS sense (i.e., adjective, adverb, verb, or noun) in each month, the aggregated word-level sentiment scores were calculated using the procedure shown in Figure 3.4.

Pairwise t-tests were conducted to show the statistical significance of the emotional differences between the two genders. As summarized in Table 3.2, the sentence-level analysis results indicate that women were significantly more subjective than men (p-value = 0.0087); the phrase-level analysis results show that women were both significantly more positive (p-value = 0.0271) and significantly more negative (p-value < 0.0001) than men. In addition, the average objective score of men was larger than that of women, but not statistically significant (p-value = 0.1935).

Table 3.2 Sentence-level and phrase-level analysis results

Sentence-level analysis result				
Comparison	**Measure**	**Women Mean/SD**	**Men Mean/SD**	**p-value**
Women < Men	Objectivity	0.136 / 0.023	0.142 / 0.026	0.1935
Women > Men	Subjectivity	0.221 / 0.023	0.209 / 0.020	0.0087**
Phrase-level analysis result				
Comparison	**Measure**	**Women Mean/SD**	**Men Mean/SD**	**p-value**
Women > Men	Positivity	0.373 / 0.052	0.350 / 0.041	0.0271*
Women > Men	Negativity	0.209 / 0.028	0.184 / 0.025	<0.0001**

Note. Significance levels * $\alpha = 0.05$ and ** $\alpha = 0.01$

Table 3.3 lists the results of the word-level analysis for different POS senses, including adjective, adverb, verb, and noun. The results of six out of all eight comparisons show that women had significantly higher polarity scores than men. Specifically, women were significantly more likely to use negative adjectives than men (p-value = 0.0004). They were also significantly more likely to use positive adverbs in comparison with men (p-value = 0.0009). In terms of verbs, women were significantly more likely to use both positive verbs (p-value = 0.0499) and negative verbs (p-value = 0.0208) compared with men. When using nouns, women were significantly more likely to use both positive nouns (p-value = 0.0229) and negative nouns (p-value = 0.0014) than men as well. However, men had a significantly higher average negative scores than women in the use of adverbs (p-value = 0.0001). The average positive score of men was also higher than that of women when using adjectives, but not statistically significant (p-value = 0.3809). Overall, the above analysis results indicate that in most cases, women tend to be more likely to express both positive and negative emotions than men.

Table 3.3 Word-level analysis results

POS	Comparison	Measure	Women Mean/SD	Men Mean/SD	p-value
Adjective	Women < Men	Positivity	0.442 / 0.024	0.444 / 0.025	0.3809
	Women > Men	Negativity	0.358 / 0.016	0.343 / 0.019	0.0004**
Adverb	Women > Men	Positivity	0.445 / 0.034	0.422 / 0.019	0.0009**
	Women < Men	Negativity	0.337 / 0.024	0.357 / 0.019	0.0001**
Verb	Women > Men	Positivity	0.347 / 0.020	0.339 / 0.013	0.0499*
	Women > Men	Negativity	0.291 / 0.013	0.286 / 0.011	0.0208*
Noun	Women > Men	Positivity	0.224 / 0.019	0.218 / 0.011	0.0229*
	Women > Men	Negativity	0.166 / 0.011	0.157 / 0.013	0.0014**

Note. Significance levels * $\alpha = 0.05$ and ** $\alpha = 0.01$

Overall, the analysis results indicate that the emotional differences between women and men in the Web forum setting are consistent with what previous research found in face-to-face settings in the physical world. That is, women are more emotionally expressive than men and they are more likely to express both positive and negative emotions.

As mentioned before, the testbed has a large number of discussions related to family issues, human rights, and political issues. I further investigated the women generated messages that convey intense emotion. I observed that many of them are discussions related to stereotyping and social roles. Forum message examples related to the two theoretical perspectives are listed in Tables 3.4-3.5, respectively. The positive and negative words and phrases detected by the PED and WED algorithms were also indicated in the message bodies (bold and italic) with another two columns showing the positive and negative ones respectively.

Stereotypes provide models of appropriate behavior which people may adopt. People who do not conform to stereotypic behavior may be punished through social rejection, whereas people who do conform may be encouraged or rewarded for their behavior in the form of social approval. The first example in Table 3.4 expresses the idea about how a man should behave with his family. It emphasizes that the appropriate behavior is responsibility toward the family, and to treat the family kindly. Both men and women have rights over each other. The second example illustrates the appropriate behavior for newly married couples. It states that a man should treat his wife "with the utmost care, consideration, and sensitivity from the very first moment" and try to "establish ties of love and affection with his wife and placate her worries and her fears about the new life she has just embarked upon, so that she feels secure and at peace with him." The next example emphasizes the importance of women's real beauty, saying that the "good character should be at the top of the list because this is where the real beauty comes from." These three examples are about the good behavior that people should demonstrate, while the other three examples caution people what they should not do related to the issues of abortion, polygyny, and family violence.

Table 3.4 Example messages related to the stereotyping perspective

Message	Positive Indicators	Negative Indicators
The husband should realize that he is a shepherd and is ***responsible*** for his flock. Allaah has ***enjoined*** upon him to treat them in a ***good*** and ***proper manner*** and to treat his family ***kindly***. Allaah says (interpretation of the meaning): "And they (women) have rights (over their husbands as regards living expenses) similar (to those of their husbands) over them (as regards ***obedience*** and ***respect***) to what is ***reasonable***, but men have a degree (of ***responsibility***) over them.	Responsible; Enjoined; Good; Proper; Manner; Kindly; Obedience; Respect; Reasonable; Responsibility	-
The first wedding night is like no other. It is the night where two people embark upon life in a whole ***new*** world with its own unique ***qualities*** and experiences. It is a doorway that is being crossed for the first time. The two people are able for the first time to ***enjoy*** what has always before been forbidden to them. This ***new*** permissibility applies to only one person. For the husband, this person is his wife, his life-partner, the woman who is going to be the mother of his children. Should not this woman ***deserve*** to be treated with the ***utmost care***, ***consideration***, and ***sensitivity*** from the very first moment? ...... The wedding night should be a night filled with ***tenderness***, intimacy, ***affection***, and ***joy***. In that night, the husband should be seeking to establish ties of ***love*** and ***affection*** with his wife and	New; Qualities; Enjoy; Deserve; Utmost; Care; Consideration; Tenderness; Affection; Joy; Love; Placate; Secure; Peace	Sensitivity; Worries; Fears

placate her ***worries*** and her ***fears*** about the ***new*** life she has just embarked upon, so that she feels ***secure*** and at ***peace*** with him.		
If you were an "***ugly***" person, how would you deal with your ***ugliness***. Would you stay at home all day, ***feeling sorry*** because you are not as ***nice*** looking as other people or would you think, look Allaah created me this way. I can't change my looks. I can only change what I have come ***control*** over like my ***character***, the way I deal with people, the level to which I ***obey*** Allaah, etc. …… There must be someone out there who will find me ***attractive*** and ***would want*** to marry me which leads to the point that ***beauty*** is in the eye of the beholder. Not all men or women go for supermodel looks because sometimes a ***good*** looking person can have an ***ego*** that is ***difficult*** to live with. ***Good character*** should be at the top of the list because this is where the ***real beauty*** comes from.	Nice; Control; Character; Obey; Attractive; Would want; Beauty; Good; Character; Real	Ugly; Ugliness; Feeling sorry; Ego; Difficult
I think that actually the ***problem*** is that many girls would rather ***die*** than have their families know they are pregnant and so they would just deal with it themselves and have an abortion. ... The scale of the ***problem*** is enormous, as is obvious from abortion statistics. I don't think much is being done at all, not for girls anyway.	-	Problem; Die
Men are allowed to marry up to 4 wives. Period. Yes, he has to do ***justice*** between them, yes, yes, yes. but that	Justice; Right	Destroying; Hurting;

comes after marriage. You risk ***destroying*** your family, relations, ***hurting*** others just to fulfill your sexual desires. And yeah you have the ***right*** to, but it doesn't mean you ***abuse*** the right.		Abuse
Beating [wife] doesn't make you manly. It makes you ***childish*** and ***violent***.	-	Childish; Violent

Table 3.5 shows some examples related to the social roles perspective that suggests that the different social roles that women and men commonly occupy lead to sex differences in social behavior. The first two examples emphasize the social role of women as child educators. Both mention that it is the mother's responsibility to teach her children, and to support homeschooling. The second example also mentions the things that girls need to learn, including "sewing, embroidery and cooking." The third example emphasizes the important role of parents in education, saying that parents must be a good example to their children. The next two examples are about women's roles in cooking. The first one states that cooking is a woman's job, while the other one states that in addition to cooking, women can do a typical man's job equally as well or even better. The last example illustrates a young woman's worries about men's role as the power center of a family.

Table 3.5 Example messages related to the social roles perspective

Message	Positive Indicators	Negative Indicators
I have a child who in two years will enter school, which I'm ***hoping*** with Allah's help will be the home in which she lives. I went to public schools all my life and I attended an Islamic school for one summer where I learned Arabic after graduating from high school. After reading a book about child ***education*** in Islam, I ***realized*** that it is my ***responsibility*** as a mother to educate my children. I have a teaching degree so I could ***teach*** my child all that she needs to ***learn***. I believe that Islamic ***education*** should be the parents ***responsibility*** and unless both parents work, they should educate their children at home. Academic subjects like ***reading***, writing, and math can also be taught at home.	Hoping; Education; Realized; Responsibility; Teach; Learn	-
I voted for homeschooling, unless the parents are ***unable*** to do so. With homeschooling, the child can get more [than] one attention, and you can also ***focus*** on other things which are ***important*** to you and your child. For example, you can incorporate various different sections of Islamic Studies, history, Quran, and ***practical*** things such as wudoo, performing salah etc. For girls, it would also be a ***good*** time to introduce ***crafts*** like sewing, ***embroidery*** and cooking. You can work with your kids to get a ***good*** balance of ***secular*** sciences, ***practical skills***, and Islamic sciences.	Focus; Important; Practical; Good; Crafts; Embroidery; Practical; Skills	Unable; Secular

A child usually grows up in the ***manner*** to which the parents make him or her ***accustomed***. During the ***blessed*** days of Ramadan, fathers and mothers have a ***great*** role to play in seizing this ***good occasion*** for their own benefit and for that of their children, and we can offer parents the following advice: ...	Manner; Accustomed; Blessed; Great; Good; Occasion	-
I don't think men should ***bother*** cooking unless they ***really really really like*** to. To be ***completely honest*** I think that cooking is a woman's job. You wouldn't wear a dress just cause its ***cute*** now would ya? I know this is ***harsh*** but ***really*** come on guys... BBQ is the ***only*** male cooking I'll ***allow*** in my household. Someone might convince me otherwise but until then men should leave the female thing to the women.	Really; Like; Completely; Honest; Allow	Bother; Harsh; Only
My mom cooks the family food out of ***love***, whenever my dad cooks (and he doesn't cook ***good***, but he think he does) he ***complains*** and ***grumps*** as if he is doing everyone a huge ***favor*** - plus he leaves a huge ***mess*** and then boasts about cooking ***better*** than my mom ever did (which is ***nowhere*** near the ***truth***, lol), but everyone is ***forced*** to agree… All in all... Some men can cook, so what? Some women can handle a job, and earn the living for her ***entire*** family, but that doesn't mean it's her ***responsibility***. Some women can equally do a ***typical*** man's job ***much better***, but they just keep quiet to let the men in their lives to feel like they are ***actually*** doing something ***beneficial***. Cause in reality, nowadays, we don't need men, at least I don't.	Love; Good; Favor; Better; Truth; Entire; Responsibility; Typical; Much better; Actually; Beneficial	Complains; Grumps; Mess; Nowhere; Forced

I think our family does need someone ***strong*** to ***rely*** on. We have no one here. Even our dad left us ***alone***. My mom says we need someone who can help me when I go to out of state for medical school. Who will go with us, for sure it won't be my dad. He always said I am worthless and won't do anything ***special*** in life. The thing I am ***worried*** about is what will happen when I will be getting married. Isn't it ***considered bad*** in Asian countries to have a step-father or divorced parents? I am ***sick and tired*** of all this. I don't want to have another dad or need any man's help at all. I wish I could just live life without marrying or dealing with these kinds of issues. ***Just devote*** my life to Allah and my career. It's just not worth it. I don't want to end up like my mom. Most Muslim women I have known are ***suffering*** like this. Even my grandmother ***died*** because my grand father used to ***abuse*** her. There was a lady in newspaper that was ***burned*** by her husband for not obeying his ***stupid*** rule. I think right now I am just full of ***worries***. I am praying these days for Allah's help. This is the place I can seek advice from, so if you can give me any ***helpful*** advice, that would be ***appreciated***. ***Please*** tell me if what my mom is saying is ***right*** or ***wrong*** because I seriously don't know.	Strong; Rely; Special; Just devote; Helpful; Appreciated; Please; Right	Alone; Worried; Considered Bad; Sick and tired; Suffering; Died; Abuse; Burned; Stupid; Worries; Wrong

3.6 Conclusions and Future Directions

3.6.1 Research Contributions

Previous studies on emotional differences between women and men in traditional, face-to-face communication find that, in general, women tend to be more emotional than men, and they are more likely to express positive emotions than men. Although some inconsistencies exist, many studies have found that women are more likely to report negative emotions as well. However, few studies have examined these differences in the new CMC media. With the advent of Web 2.0, a large and increasing amount of user-generated content has been emerging in cyberspace. People have begun to use different Web 2.0 platforms to exchange information and express opinions. Text-based, asynchronous CMC has become popular, because it allows people from different places to communicate with each other at different times in a flexible manner.

To examine the emotional differences between the two genders in the new communication media, an automatic emotion detection framework using sentiment analysis techniques was proposed in this study. Unlike most previous research, the proposed framework takes into account the longitudinal textual data from an entire text-based CMC site to conduct the emotional difference analysis automatically, which can provide more comprehensive and unbiased analysis results. Specifically, different algorithms were developed to analyze the sentence-level subjectivity and phrase- and word-level polarity, and then conducted statistical analysis to compare the differences between women and men. To the best of my knowledge, this is the first study to utilize advanced sentiment analysis techniques to analyze emotional differences between the two genders.

In addition, no previous studies have specifically examined these differences in Web forum communication. Using the proposed framework, an empirical experiment was conducted on a large and long-standing international women's political forum with more than 30,000 messages. The sentence-level analysis results indicated that women were significantly more subjective than men. The phrase- and word-level analysis results showed that in general women were significantly more likely to express both positive and negative emotions as compared to men. By investigating the emotional content generated by women, the existence of discussions related to stereotyping and social roles was observed.

3.6.2 Future Research Directions

The present study has some limitations that may be addressed in future research. First, as summarized by Mauss and Robison (2009), the measurement of emotion with any instrument is not perfect and can have potential biases since human perception is inherently variable (Fox 2008; Thelwall et al. 2010). Future studies can adopt other methodologies to examine the emotional differences between women and men to see whether consistent results can be found. Second, the types of emotions measured in this study focus on positivity versus negativity. Future research can examine more specific types of emotions, such as happiness, love, life satisfaction, etc., versus fear, sadness, anger, etc. Third, in this study, an empirical experiment was conducted on Web forum content; future research can adopt the proposed generic framework to examine the emotional differences between women and men in other text-based CMC, such as email, blogs, and wikis. Moreover, although an English-language testbed was used in this study, the proposed framework could be applied to other languages, and a multilingual emotion analysis component could be developed to support such research.

CHAPTER 4. THE IMPACTS OF VIRTUAL GENDER, VIRTUAL AGE, AND REGION THEME ON AVATAR ACTIVITY IN THE VIRTUAL WORLD

4.1 Introduction

Chapters 2 and 3 focused on examining the gender differences in the online world. Chapter 2 developed a research framework to classify the two genders based on their writing styles and topics of interest. Chapter 3 further examined the emotional differences between the two genders by developing a research framework using sentiment analysis techniques. In this Chapter, gender and age differences are examined in the virtual world. Specifically, guided by the theories of social presence, social role, and gender role, the main effects and interaction effects of avatar virtual gender, virtual age, and region theme on avatars' physical activities in the virtual world are examined. To enable the analyses, an integrated technical framework for avatar data collection is developed by combining an improved bot-based approach and a spider-based approach.

Virtual worlds are rapidly becoming an important place in people's lives, in addition to the real world where they physically live (Chen 2009). It has been reported that millions of people from different areas of the world spend on average 22 hours a week in virtual worlds interacting with others through their avatars (Yee and Bailenson 2007). The Gartner Group has estimated that 80 percent of active Internet users will have a "second life" in the virtual world by the end of 2011 (www.gartner.com/it/page.jsp?id=503861). Second Life (http://secondlife.com), developed by Linden Lab and publicly launched in 2003, is currently the most popular

3D virtual world where users (or residents) interact with each other through customized avatars. Since the site's inception, the total number of Second Life residents has grown dramatically—from 2.2 million residents in December 2006 to 8.3 million in August 2007 (Messinger et al. 2009). Even U.S. President Barack Obama and Secretary of State Hillary Clinton had a presence in Second Life during their 2008 campaigns (Chen 2009).

Internet and Web technology advances enable users to actively participate in the Internet instead of only passively acquiring information. The amount of user-generated social media data is growing tremendously (Wang et al. 2007). Virtual worlds are rich in this data, of interest to researchers in areas such as business, social science, communication, politics, and education (Chen 2009; Hendaoui et al. 2008; Messinger et al. 2009). For example, virtual worlds can offer untapped marketing potential because of their ability to generate sustained consumer engagement with a brand (Hemp 2006). Virtual worlds can present information and knowledge vividly, which makes them an ideal medium for e-learning (Franceschi et al. 2009; Hassell et al. 2009). It has been expected that the emerging virtual world market could reach billions of dollars in the coming years (Hendaoui et al. 2008).

In addition, virtual worlds can offer many potential benefits to researchers to conduct their studies. For example, compared to formal experiments in the physical world, researchers can easily recruit a much larger number of subjects in virtual worlds with a facility comparable to the physical world (Bainbridge 2007). It is also possible to study social and ethical issues or other issues that cannot be performed in the physical world, for example transmitted diseases can be studied by manipulating the appearance and behavior of avatars (Bainbridge 2007; Gordon et al. 2009).

In order to fully use the benefits of virtual worlds, it is important to know whether people display consistent behavior across the virtual and physical worlds. If so, it then becomes possible to test certain hypotheses and phenomena that cannot be easily tested in the physical world and generalize the results to the physical world to help further develop guidelines for people's real lives (Bainbridge 2007; Yee et al. 2007). This can lead to great benefits in various aspects of people's lives. However, little effort has been made to specifically examine whether or not people display consistent behavior across both types of worlds. One major obstacle is data collection. As this new media still in its development stage, few virtual worlds provide easy data-collection functions.

In this research, an exploratory study was conducted to examine whether real-life social norms hold in the virtual world. Specifically, the impacts of avatar gender, avatar age, and region theme factors, as well as their interaction effects on avatar activities were examined. In addition, a systematic framework was developed to collect avatar-related data using a combination of bot- and spider-based approaches. Following the framework, an experimental study was conducted on Second Life.

The remainder of this chapter is organized as follows. Section 4.2 lists the research questions of this study. Section 4.3 provides the theoretical background. Section 4.4 describes the proposed research model. Section 4.5 shows detailed information of the research method, including the improved bot-based approach and the spider-based approach. Section 4.6 discusses data analyses and results. Section 4.7 summarizes the contributions of this study, the implications, and future research directions.

4.2 Research Questions

As mentioned above, the fundamental research question of this study is:

Do people display behavior that is largely consistent across the virtual world and the physical world?

Because this question is so broad, trying to answer it is far beyond the scope of a single research study.

Therefore, in this study, I focus on two more specific research questions:

1. Do real-life social norms hold in the virtual world?
2. What are the major factors that can influence people's behavior in the virtual world?

To answer the above research questions, Section 4.3 provides the related theoretical background. Sections 4.4-4.6 present the research model, method, and the data analysis results respectively.

4.3 Theoretical Background

4.3.1 Social Presence Theory

As an important theory related to communication media, social presence theory (Short et al. 1976) measures communication media based on the degree of awareness of the other person in a communication interaction. According to social presence theory (Short et al. 1976), using communication medium with appropriate social presence can lead to effective communication. In general, a communication medium with higher social presence can help achieve better understanding among speakers. The social presence level of a communication medium can be changed by adding or removing certain communication modality such as verbal and non-verbal cues, and immediate feedback exchange (Sallnas et al. 2000). Face-to-face communication has been considered to have the most social presence, while text-based communication has the least social presence.

For computer-mediated communication (CMC), social presence refers to how people represent themselves in the new media (Kehrwald 2008). For example, user ID can be the representation of an online forum user in the online world, and an avatar is the representation of a person in the virtual world. The immersive environments of virtual worlds can increase the perceptive experience of individuals, typically via the visual channel (Harris et al. 2009). Such immersion can lead to an increased feeling of social presence (Harris et al. 2009). Virtual worlds substantially increase the sense of actual presence in CMC. They blur the distinction between the face-to-face communication in the physical world and CMC in online and virtual worlds by providing the bandwidth for transmitting many types of signals in face-to-face communications (Blascovich et al. 2002).

The immersive nature offers virtual worlds higher social presence compared to other CMC media (Blascovich et al. 2002; Harris et al. 2009). Signals in virtual worlds can be conveyed via both verbal and nonverbal communication channels. In addition to the presence of others, virtual worlds also enable the presence in terms of personal presence and environmental presence. Virtual worlds offer the greatest sense of actual presence and convey important cues. Social interactions in virtual worlds are expected to most closely resemble face-to-face interactions in the physical world.

Because of the immersive nature of virtual worlds, previous research has suggested leveraging them as a platform to examine how virtual avatars can influence realistic user behavior (Blascovich et al. 2002). Increased social presence in virtual environments can increase the likelihood that individuals will respond in realistic ways as in the physical world (Harris et al. 2009). Therefore, it could be expected that social behavior and norms in virtual worlds are comparable to those in the physical world (Blascovich et al. 2002; Yee et al. 2007).

4.3.2 Social Role Theory

Social role theory (Biddle 1986) suggests that people act according to expectations from themselves and others. Different roles (such as parent or lawyer) that people perform in their daily lives have different expectations. People’s activity is generally consistent with their role expectations (Hindin 2007).

Role behavior is influenced by role expectations for appropriate behavior in that position (Thompson 2001). Each person typically has more than one role. Each role consists of a set of rules, rights, duties, and norms that function as plans or blueprints to guide behavior. Social role theory is predictive (Biddle 1986). From the expectations for a particular role, the activity of a person who performs that role can be predicted.

4.3.3 Gender Role Theory

Gender role theory (Eagly and Karau 1991) states that, in general, gender differences in behavior and personality are socially constructed. Gender roles are considered as a set of social and behavioral norms appropriate for individuals of a specific gender.

Traditionally, women are more likely than men to fill caretaker roles in the home (e.g., wife and mother) and when employed for pay (e.g., teacher and nurse); men, however, are more likely than women to hold jobs in providing the material resources in paid employment settings (Fischer and Manstead 2004; Grossman and Wood 1993). As argued by Chodorow (1994), girls tend to treat their mothers as role models and follow their behavior to be more emotional, social, and caring about others; however, boys are encouraged to be brave and independent.

Gender role theory suggests that "individuals internalize cultural expectations about their gender because social pressures external to the individual favor behavior consistent with their prescribed gender role" (Kidder 2002, p. 630). People use these expectations to guide their behavior, such as physical activities (Malcom 2003) and emotions (Palapattu et al. 2006). In two recently reported studies on gender differences in physical activity, researchers found that females of all ages (from school children to over 70s) are less active than their male peers (BBC 2009; ScienceDaily 2009). In addition, gender roles have been found to be moderated by age (Malcom 2003). As two major markers of social identities, gender and age are important social roles to consider when examining people's behavior differences (Blascovich et al. 2002). Virtual worlds make it quite possible and relatively easy to manipulate these two types of roles experimentally (Blascovich et al. 2002).

For avatar gender differences in virtual worlds, previous research has shown that the two genders have social space differences. Male avatars tend to keep larger interpersonal distances with other male avatars compare to the interpersonal distances between two female avatars (Yee et al. 2007). However, no study has specifically examined the impact of avatar gender on avatars' physical activities. For virtual age differences, Harris et al. (2009) conducted a lab experiment with 80 students and found that in general avatar activity declines with increased use experience. The major limitation of their study is the relatively small sample size and the use of student subjects instead of general Second Life residents. In addition, previous research has suggested taking context into account for future studies on virtual worlds (Yee et al. 2007). However, no existing study has empirically investigated avatar behavior differences in different contexts such as different types of regions.

Motivated by the above discussion, this study investigated the impacts of avatars' virtual gender, virtual age, and region theme, as well as their interactions, on avatars' physical activities.

4.4 Research Model

Figure 4.1 shows the proposed research model. The dependent variable is avatar activity which is defined at two levels: high-active and low-active. The independent variables include virtual gender, virtual age, and region theme.

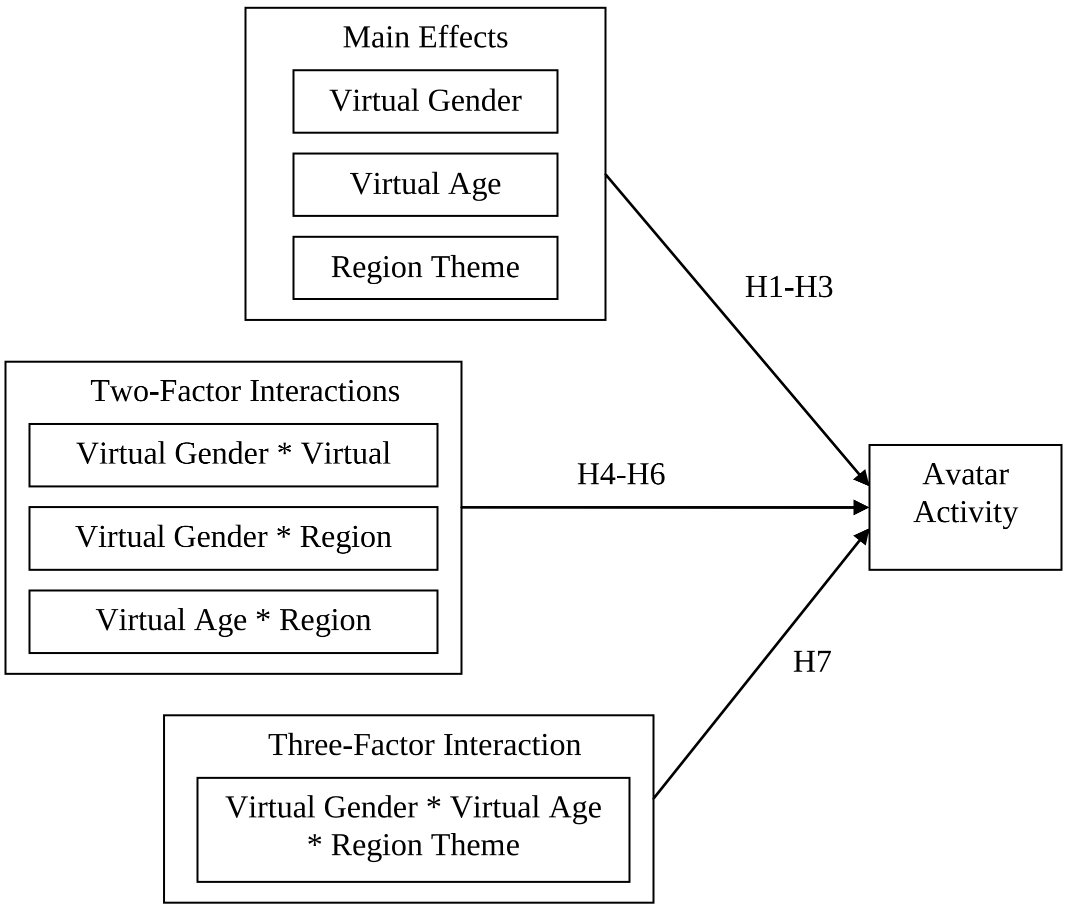

Figure 4.1 Research model

Virtual gender is the gender of a given avatar. In virtual worlds, in addition to females and males, users may also be able to set their avatars to be non-humans (Yee et al. 2007). This study focuses only on the two genders, females and males, which are the dominated cases in real lives. Similar to a person's age in the physical world, virtual age of an avatar refers to how long the avatar has existed in a virtual world. In this study, virtual age is examined on two levels: young and old.

As in the physical world, there are different places where people can visit according to their different needs in everyday lives. Similarly, in virtual worlds, there are different regions that people can visit for different purposes. Different from the physical world, as a complex platform with technical support, virtual worlds often impose a considerable learning curve on users, especially for creating virtual items (Diversified Media Design et al. 2007; Messinger et al. 2009). Therefore, help-supporting regions are often very popular in virtual worlds. Different from help-supporting regions which are always non-profit and based on altruism, another type of popular region in virtual worlds is based on commercial transactions, that is, buying virtual objects such as virtual clothes for avatars (Messinger et al. 2009). This study focuses on these two types of regions: help-supporting regions and commercial transaction regions.

According to social role theory, human behavior is guided by expectations associated with the roles people have (Biddle 1986; Thompson 2001). Such roles can be the gender, age, or certain positions of a given person. According to gender role theory (Eagly and Karau 1991; Kidder 2002), females and males behave differently since cultural expectations of the two genders are different and social pressures guide individuals to follow the proper behavior consistent with their prescribed gender role. As another important social role, age is also expected to influence people's behavior (Blascovich et al. 2002). Influenced by the associated social norms, people in different positions or at different locations are also expected to behave differently (Gordon-Larsen et al. 2000; Messinger et al. 2009).

According to social presence theory, their immersive nature offers virtual worlds higher social presence compared to other CMC media (Harris et al. 2009). The rich communication channels of virtual worlds blur the distinction between actual, face-to-face communication and virtual world communication (Blascovich et al. 2002). Because of the high realism of immersion in virtual worlds, it is expected that social behavior and norms in virtual worlds are comparable to those in the real, physical world (Blascovich et al. 2002; Yee et al. 2007).

In the physical world, in general, males are often more active than females (BBC 2009; Belcher et al. 2010; ScienceDaily 2009; Trost et al. 2002). It is commonly agreed that young people are often more active than old people (Belcher et al. 2010; Caspersen et al. 2000; Sallis 2000; Trost et al. 2002). Therefore, I posit that in virtual worlds male avatars are more active than female avatars and avatars of younger ages are more active than the ones with relatively older ages.

H1: Male avatars are more active than female avatars. (M > F)

H2: Young-aged avatars are more active than old-aged avatars. (Y > O)

Previous research has shown location to be an important factor in influencing people's physical activities in the physical world (Gordon-Larsen et al. 2000). In virtual worlds, it is argued that context (i.e., different region themes) might play an important role in understanding avatars' virtual behavior (Yee et al. 2007). However, few existing virtual world studies have considered the context factor. This study uses region theme as the context variable and examines its relationship with avatar activity and interactions with other independent variables.

Region theme is expected to influence avatar activity since avatars in different types of regions could be influenced by different social roles associated with the themes of these regions. Based on the observations, in virtual worlds, when people are seeking or offering help, they typically talk to each other using the verbal channels and little physical action is needed from the avatars. In contrast, in order to purchase virtual items in commercial transition regions, avatars need to walk around in virtual shopping centers or malls to look for and select the items they are interested in. I thus posit that avatars in commercial transition regions are more active compared to avatars in help-supporting regions.

> H3: Avatars in commercial transaction regions are more active than avatars in help-supporting regions. (C > H)

Gender roles have been found to be moderated by the age factor (Malcom 2003). I therefore expect significant interaction effects between virtual gender and virtual age (virtual gender * virtual age) when influencing avatar activity in virtual worlds.

> H4: Male avatars with young (old) virtual age are more active than female avatars with young (old) virtual age (MY > FY and MO > FO), while male (female) avatars with young virtual age are more active than male (female) avatars with old virtual age (MY > MO and FY > FO).

I also expect that region theme has significant interaction effects with virtual gender (virtual gender * region theme) and virtual age (virtual age * region theme) respectively when influencing avatar activity in virtual worlds.

> H5: Male avatars in commercial transaction (help-supporting) regions are more active than female avatars in commercial transaction (help-supporting) regions (MC > FC and MH > FH), while male (female) avatars in commercial transaction regions are more active than male (female) avatars in help-supporting regions. (MC > MH and FC > FH).

> H6: Young-aged avatars in commercial transaction (help-supporting) regions are more active than old-aged avatars in commercial transaction (help-supporting) regions (YC > OC and YH > OH), while young-aged (old-aged) avatars in commercial transaction regions are more active than young-aged (old-aged) avatars in help-supporting regions (YC > YH and OC > OH).

A significant three-factor interaction effect among virtual gender, virtual age, and region theme (virtual gender * virtual age * region theme) is also expected when influencing avatar activity in virtual worlds.

> H7: Male avatars are more active than female avatars in all four conditions of virtual age * region theme (MYC > FYC, MOC > FOC, MYH > FYH, and MOH > FOH); young-aged avatars are more active than old-aged avatars in all four conditions of virtual gender * region theme (MYC > MOC, FYC > FOC, MYH > MOH, and FYH > FOH); avatars in commercial transaction regions are more active than avatars in help-supporting regions in all four conditions of virtual gender * virtual age (MYC > MYH, FYC > FYH, MOC > MOH, and FOC > FOH).

4.5 Research Method

Collecting avatar behavior and profile data from virtual worlds is critical yet difficult. Most previous studies used self-reported questionnaires to collect this data. The major limitations of this method include the unreliability of self-reported measures and the relatively small sample size that can be collected (Yee and Bailenson 2008). Other studies have used the Linden Scripting Language (LSL) to automatically or semi-automatically collect avatar-related data (Yee and Bailenson 2008).

Second Life users use LSL to control their avatars' behavior. LSL is a state-event-driven scripting language. A script can contain variables, functions, and states. Each state contains a description of how to react to events (http://wiki.secondlife.com/wiki). Currently, little research has been done on avatar data collection using LSL. Yee's group, which is well known for this type of research, developed a semi-automatic, bot-based approach using LSL (Yee and Bailenson 2008; Yee et al. 2007).

Although quite advanced, their approach has some limitations due to the restrictions of LSL. For example, a bot can collect only a relatively small number of records about avatars close to them; therefore, human intervention is needed to manually move the bot close to avatars in order to collect their behavioral data. Because bots are attached to avatars, when the avatars change their outfits, their attached bots automatically detach as well. Furthermore, the method cannot collect avatar profile information. To determine whether an avatar was male or female, experimenters had to stay in Second Life and watch and write down the gender of each encountered avatar (Yee et al. 2007).

To address these limitations, this study developed an improved bot-based approach that uses LibOpenMetaverse (www.libsecondlife.org/projects/libopenmetaverse), a software library that can be used as a third-party application to communicate with the Second Life server. LibOpenMetaverse was developed using C#, a flexible object-oriented programming language. Although both LSL and LibOpenMetaverse provide a similar set of functions for data collection and avatar manipulation, LibOpenMetaverse allows each bot to conduct a much larger-scale data collection. Table 4.1 lists the advantages of LibOpenMetaverse compared to LSL in terms of collecting avatar behavioral data.

Table 4.1 Advantages of LibOpenMetaverse over LSL for collecting avatar behavioral data

LibOpenMetaverse	**Linden Scripting Language (LSL)**
Developed based on C#, a flexible object-oriented language	Developed based on an inflexible state-event based language
Programs are compiled locally	Programs are compiled in the Second Life server; speed depends on the server traffic
Runs in command lines; can run multiple threads on one machine	Needs to open the Second Life viewer, which requires a high-quality graphics card; typically can only run one Second Life viewer on one machine
Avatars can be logged on to the Second Life server through LibOpenMetaverse clients and act as bots themselves	Can be used to create object-like bots and attach them to avatars to collect data
No memory limit on a bot	Only allocates up to 16 Kbytes of memory for each bot; cannot collect a large amount of data at the same time

Because LibOpenMetaverse programs are compiled locally and run in command lines, multiple threads can be opened to collect data much faster without being affected by Second Life server's traffic. Another major advantage of LibOpenMetaverse is that it does not limit the amount of memory available to each bot, whereas LSL allocates only 16 Kbytes of memory for each bot. Furthermore, because the bot-based approach cannot collect avatar profile data, a spider-based approach is used to obtain that information. By combining an improved bot-based approach developed using LibOpenMetaverse with a spider-based approach, the integrated framework can collect both avatar behavioral and profile data.

Figure 4.2 shows the proposed integrated framework. The improved bot-based approach is used to collect avatar behavioral data, and the spider-based approach is used to collect avatar profile data (such as virtual gender, virtual age, etc.).

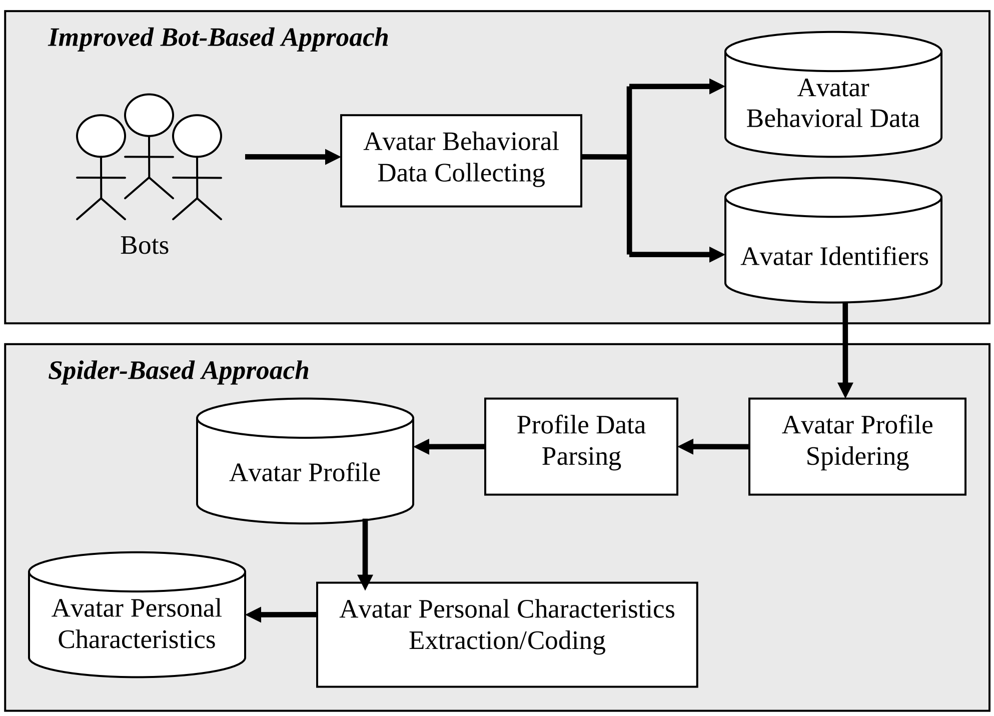

Figure 4.2 The integrated framework for avatar data collection

4.5.1 Improved Bot-Based Approach

By logging into the Second Life server using LibOpenMetaverse as clients, bots appear as real avatars and automatically collect other avatars' behavioral data in a given region. LibOpenMetaverse version 0.8-Pre, the newest version currently available (http://jira.openmv.org/browse/LIBOMV), was used. One bot is sent to a given region so it can collect both the behavioral data and the avatar identifiers (such as names) at the same time. In Second Life, a region comprises an area of 65536 m^2 (16.1943 acres), being 256 meters on each side. Using LibOpenMetaverse, one bot is able to cover the entire area of one region. Avatar names are needed for the subsequent spider-based approach to collect avatars' profile data. In this way, a given avatar's profile information can be matched to his or her behavior.

The bots collect data in real time, sampling it every five seconds. That is, for each avatar in a given region, one record is collected about his or her ongoing action every five seconds. As previous studies suggested (Harris et al. 2009; Yee and Bailenson 2008), a 5- or 10-second sampling resolution is sufficient for collecting a substantial amount of data while keeping a reasonably low workload on the Second Life server. An avatar can conduct more than one action at the same time, such as running while turning left. In this case, both actions in one record are captured. To minimize the potential biases introduced by the bots, during the data collection process, the bots neither move nor communicate with others in the regions. A record includes avatar ID, avatar name, avatar's group in

Second Life, region name, avatar position, avatar face angle, action ID, action name, and time stamp. Figure 4.3 shows an example of the records collected by a bot.

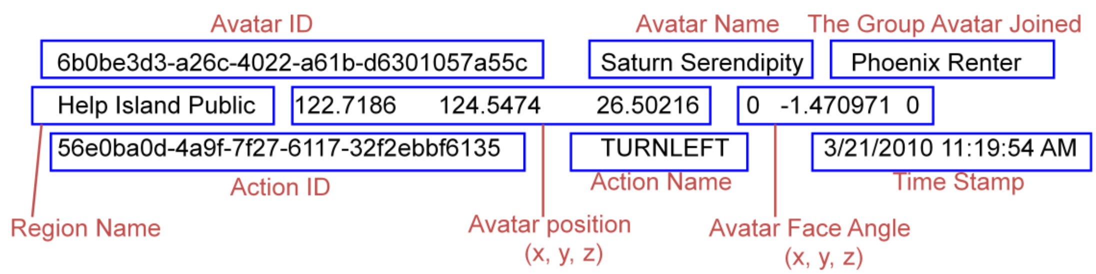

Figure 4.3 An example of the records collected by the bots

Figure 4.4 shows the pseudocode of the improved bot-based approach. Bots are first initialized and sent to the designated regions, and then they collect the avatar behavioral data. A detected avatar must first be checked to see if it is a bot before data is collected. If so, it is ignored.

Initialize bots and send them to the designated regions
For each bot *b*
    Logon to the Second Life server using the LibOpenMetaverse client with parameters (avatar name, password, region name)

Collect avatar behavioral data
For every 5 seconds
    b requests avatar properties
    If *Avatar properties response received == TRUE*
        For each avatar *a* in the region
            If *a* == *b*
                Ignore
            Else
                Add a new record containing avatar identifiers and the current action(s)
Logoff *b* from the Second Life server

Figure 4.4 Pseudocode for the improved bot-based approach

4.5.2 Spider-Based Approach

As Figure 4.2 shows, the three components of the spider-based approach are avatar profile spidering, profile data parsing, and avatar personal characteristics extraction and coding.

The Second Life search engine (http://search.secondlife.com) provides a function to search for a given avatar's detailed information by providing the avatar's name as input. An automatic spidering program was developed to collect the profile pages of all the avatars whose behavioral information was collected. The spidering program sends the names of these avatars to the Second Life search engine, one at a time, and the search engine returns the detailed profile pages in HTML format. Because the search engine uses fuzzy matching for the avatar names, the program checks whether a returned page is for exactly the same avatar name as input. If so, the page is kept; otherwise, it is excluded. Figure 4.5 shows an example of the collected avatar profile pages.

Figure 4.5 An example of the avatar profile pages

Data fields in the collected profile pages include an avatar's name, birthday, picture, and short self-introduction. An automatic parsing program parses these data fields out and stores them in a relational database. Avatar birthday information can be used to calculate an avatar's virtual age, which refers to how long the avatar has existed in its virtual world.

Some important avatar personal characteristics (such as gender) are not included in the avatar profile page. To collect such information, the data fields collected in the profile data parsing component—avatar names, self-introductions, and profile pictures—are utilized. Although this task is still done manually, researchers do not need to stay in

Second Life and can do it more efficiently offline. Figure 4.6 shows the pseudocode of the three components in the spider-based approach used to collect avatar profile data.

Avatar profile spidering
For each avatar *a*
  Send the first name and family name of *a* to the Second Life search engine
  Get a list of returned avatars $a_1, a_2, \ldots, a_n$ (the search engine uses fuzzy matching)
  For each a_i
    If a_i . *first name* == *a* . *first name AND* a_i . *family name* == *a* . *family name*
      Spider the detailed profile page of a_i
    Else
      Ignore

Profile data parsing
For each detailed avatar profile page
  Parse out the avatar birthday, self introduction, and avatar profile picture

Avatar personal characteristics extraction/coding
For each avatar *a*
  Generate a file including (avatar name, self introduction, profile picture)
  If *a's name OR a's self introduction OR a's profile picture* can indicate a given personal characteristic (e.g., gender) of *a*
    Assign the indicated characteristics to *a*
  Else
    Label the characteristics of *a* as unknown

Figure 4.6 Pseudocode for the spider-based approach

4.5.3 Sample Size and Measures

As mentioned in Section 4.4, this study focuses on two types of regions: help-supporting regions (H) and commercial transaction regions (C). Three regions were used in the experiment, one commercial transaction region and two help-supporting regions.

The commercial transaction region is called “Exchange” and is one of the most popular regions in Xstreet SL Marketplace. To achieve a relatively balanced number of avatars compared to the “Exchange” region, two help-supporting regions were included: “Help Island Public” and “Hyannisport.” Both are popular regions where people help each other to learn to use various characteristics of Second Life. Three bots were sent to and stayed in these regions, one in each, for a week, collecting a total of 527,311 avatar behavioral records using five-second sampling.

Virtual gender is defined at two levels: male (M) and female (F). After parsing out the avatar names, self-introductions, and profile pictures, three coders were asked to determine gender. Only the avatars whose gender could clearly be identified as either male or female without any disagreement or confusion among all three coders were kept. Otherwise, they were excluded from the data collection. The goal was to collect 500 avatars with clear gender information identified for each gender and ended up with 537 male avatars and 797 female avatars.

Virtual age is defined at two levels: young (Y) and old (O). For each avatar, the specific virtual age was calculated based on his or her birthday information. Because no study has yet provided a guideline for the cutoff between young and old avatars, an exploratory analysis was conducted and found that three months was a reasonably good cutoff to achieve relatively balanced numbers for the two age groups. In addition, based on my own and my colleagues’ experiences as Second Life users, three months is a reasonable time to pass through the technology learning curve associated with Second

Life. Table 4.2 lists the number of avatars broken down by virtual gender, virtual age, and region theme.

Table 4.2 Number of avatars (subject sample size)

Factor	**Type**	**Number of Avatars**
Virtual Gender	M	537
	F	797
Virtual Age	Y	475
	O	859
Region Theme	C	595
	H	739
Total		*1,334*

For avatars' physical actions, The LibOpenMetaverse library can capture many types of avatar actions in Second Life, but an exploration revealed that avatars tend to perform only a relatively small set of physical actions. In this study, the most-widely used actions, including flying, hovering, running, sitting, standing, striding, turning left, turning right, and walking were included. Among them, flying, hovering, running, striding, turning left, turning right, and walking were considered to be high-active actions, while sitting and standing considered to be low-active actions. The percentage of times each avatar performed the high-active (low-active) actions divided by the total number of records collected was calculated to measure the avatar activity.

4.6 Data Analysis and Results

ANOVA and independent group T-test analyses were conducted to test the hypotheses. Tables 4.3 and 4.4 show the main effect testing results for high-active and low-active actions respectively.

Table 4.3 Main effect testing results for high-active actions (H1-H3)

High-Activity	**N**	**Mean**	**Std. Dev.**	**F**	**p-value**
Virtual Gender				13.77	0.0002
Male (M) (t = 3.70, p-value = 0.0001)	537	0.47	0.26		
Female (F)	797	0.41	0.26		
Virtual Age				38.29	<0.0001
Young (Y) (t = 6.16, p-value < 0.0001)	475	0.49	0.26		
Old (O)	859	0.40	0.25		
Region Theme				70.09	<0.0001
Commercial Transactions (C) (t = 8.37, p-value < 0.0001)	595	0.50	0.28		
Help-supporting (H)	739	0.38	0.23		

Note. All significant at 0.001 level.

Table 4.4 Main effect testing results for low-active actions (H1-H3)

Low-Activity	**N**	**Mean**	**Std. Dev.**	**F**	**p-value**
Virtual Gender				6.68	0.0098
Male (M) (t = 2.58, p-value = 0.0051)	537	0.79	0.15		
Female (F)	797	0.81	0.15		
Virtual Age				31.19	<0.0001
Young (Y) (t = 5.47, p-value < 0.0001)	475	0.77	0.16		
Old (O)	859	0.82	0.15		
Region Theme				22.74	<0.0001
Commercial Transaction (C) (t = 4.77, p-value < 0.0001)	595	0.78	0.16		
Help-support (H)	739	0.82	0.14		

Note. All significant at 0.001 level.

For both high-active and low-active actions, the main effects of virtual gender (F = 13.77, p-value = 0.0002 and F = 6.68, p-value = 0.0098), virtual age (F = 38.29, p-value < 0.0001 and F = 31.19, p-value < 0.0001), and region theme (F = 70.09, p-value < 0.0001 and F = 22.74, p-value < 0.0001) were all significant.

For detailed t-tests, male avatars were significantly more likely to perform high-active actions than female avatars (t = 3.70, p-value = 0.0001). In contrast, female avatars were significantly more likely to perform low-active actions than male avatars (t = 2.58, p-value = 0.0051). In addition, young-aged avatars were significantly more likely to perform high-active actions than old-aged avatars (t = 6.16, p-value < 0.0001). In contrast, old-aged avatars were significantly more likely to perform low-active actions than young-aged avatars (t = 5.47, p-value < 0.0001). Avatars in the commercial transition region were significantly more likely to perform high-active actions than avatars in the help-supporting regions (t = 8.37, p-value < 0.0001). In contrast, avatars in the help-supporting regions were significantly more likely to perform low-active actions than avatars in the commercial transition region (t = 4.77, p-value < 0.0001). Therefore, H1-H3 were all supported.

Tables 4.5 and 4.6 show the two-factor interaction effect testing results for high-active and low-active actions respectively.

Table 4.5 Two-factor interaction effect testing results for high-active actions (H4-H6)

High-Activity	**N**	**Mean**	**Std. Dev.**	**F**	**p-value**
*Virtual Gender * Virtual Age*				16.34	<0.0001
M + Y (MY) (MY vs. FY: t = 1.85, p-value = 0.0322)	218	0.52	0.25		
F + Y (FY) (MO vs. FO: t = 2.63, p-value = 0.0044)	257	0.47	0.26		
M + O (MO) (MY vs. MO: t = 3.76, p-value = 0.0001)	319	0.43	0.26		
F + O (FO) (FY vs. FO: t = 4.50, p-value < 0.0001)	540	0.39	0.25		
				31.40	<0.0001
*Virtual Gender * Region Theme*					
M + C (MC) (MC vs. FC: t = 2.66, p-value = 0.0041)	203	0.54	0.29		
F + C (FC) (MH vs. FH: t = 4.16, p-value < 0.0001)	392	0.48	0.28		
M + H (MH) (MC vs. MH: t = 5.39, p-value < 0.0001)	334	0.42	0.23		
F + H (FH) (FC vs. FH: t = 7.12, p-value < 0.0001)	405	0.35	0.22		
*Virtual Age * Region Theme*				49.89	<0.0001
Y + C (YC) (YC vs. OC: t = 2.23, p-value = 0.0132)	154	0.54	0.29		
O + C (CO) (YH vs. OH: t = 9.55, p-value < 0.0001)	441	0.48	0.28		
Y + H (YH) (YC vs. YH: t = 3.03, p-value = 0.0013)	321	0.47	0.23		
O + H (OH) (OC vs. OH: t = 10.15, p-value <0.0001)	418	0.32	0.20		

Note. All significant at 0.05 level.

Table 4.6 Two-factor interaction effect testing results for low-active actions (H4-H6)

Low-Activity	**N**	**Mean**	**Std. Dev.**	**F**	**p-value**
*Virtual Gender * Virtual Age*				11.99	<0.0001
M + Y (MY) (MY vs. FY: t = 0.96, p-value = 0.1676)*	218	0.76	0.16		
F + Y (FY) (MO vs. FO: t = 1.97, p-value = 0.0247)	257	0.77	0.16		
M + O (MO) (MY vs. MO: t = 3.16, p-value < 0.0001)	319	0.80	0.15		
F + O (FO) (FY vs. FO: t = 4.24, p-value < 0.0001)	540	0.82	0.15		
*Virtual Gender * Region Theme*				10.97	<0.0001
M + C (MC) (MC vs. FC: t = 1.99, p-value = 0.0235)	203	0.76	0.16		
F + C (FC) (MH vs. FH: t = 2.47, p-value = 0.0069)	392	0.79	0.16		
M + H (MH) (MC vs. MH: t = 3.14, p-value = 0.0009)	334	0.80	0.15		
F + H (FH) (FC vs. FH: t = 3.97, p-value < 0.0001)	405	0.83	0.14		
*Virtual Age * Region Theme*				23.94	<0.0001
Y + C (YC) (YC vs. OC: t = 2.26, p-value = 0.0124)	154	0.75	0.17		
O + C (CO) (YH vs. OH: t = 6.90, p-value < 0.0001)	441	0.79	0.15		
Y + H (YH) (YC vs. YH: t = 1.69, p-value = 0.0462)	321	0.78	0.15		
O + H (OH) (OC vs. OH: t = 6.26, p-value < 0.0001)	418	0.85	0.13		

Note. * Not significant. All the others are significant at 0.05 level.

For both high-active and low-active actions, the two-factor interaction effects of virtual gender * virtual age (F = 16.34, p-value < 0.0001 and F = 11.99, p-value < 0.0001), virtual gender * region theme (F = 31.40, p-value < 0.0001 and F = 10.97, p-value < 0.0001), and virtual age * region theme (F = 49.89, p-value < 0.0001 and F = 23.94, p-value < 0.0001) were all significant.

For the detailed t-test results about virtual gender * virtual age, as to each age group, male avatars were significantly more likely to perform high-active actions than female avatars ($t = 1.85$, p-value $= 0.0322$ for young-aged avatars; $t = 2.63$, p-value $= 0.0044$ for old-aged avatars), while female avatars were more likely to perform low-active actions than male avatars. This was statistically significant for the old-aged group ($t = 1.97$, p-value $= 0.0247$) but not for the young-aged group ($t = 0.96$, p-value $= 0.1676$). As to each gender group, young-aged avatars were significantly more likely to perform high-active actions than old-aged avatars ($t = 3.76$, p-value $= 0.0001$ for male avatars; $t = 4.50$, p-value < 0.0001 for female avatars), while old-aged avatars were significantly more likely to perform low-active actions than young-aged avatars ($t = 3.16$, p-value < 0.0001 for male avatars; $t = 4.24$, p-value < 0.0001 for female avatars). Therefore, H4 was supported in most cases except for the comparison between the two gender groups with young-aged avatars on low-active actions.

For detailed t-test results about virtual gender * region theme, as to each region theme, male avatars were significantly more likely to perform high-active actions than female avatars ($t = 2.66$, p-value $= 0.0041$ for the commercial transaction region; $t = 4.16$, p-value < 0.0001 for the help-supporting regions), while female avatars were significantly more likely to perform low-active actions than male avatars ($t = 1.99$, p-value $= 0.0235$ for the commercial transaction region; $t = 2.47$, p-value $= 0.0069$ for the help-supporting regions). As to each gender group, avatars in the commercial transaction region were significantly more likely to perform high-active actions than avatars in the help-supporting regions ($t = 5.39$, p-value < 0.0001 for male avatars; $t = 7.12$, p-value $<$

0.0001 for female avatars), while avatars in the help-supporting regions were significantly more likely to perform low-active actions than avatars in the commercial transaction region (t = 3.14, p-value = 0.0009 for male avatars; t = 3.97, p-value < 0.0001 for female avatars). Therefore, H5 was supported in all cases.

For detailed t-test results about virtual age * region theme, as to each region theme, young-aged avatars were significantly more likely to perform high-active actions than old-aged avatars (t = 2.23, p-value = 0.0132 for the commercial transaction region; t = 9.55, p-value < 0.0001 for the help-supporting regions), while old-aged avatars were significantly more likely to perform low-active actions than young-aged avatars (t = 2.26, p-value = 0.0124 for the commercial transaction region; t = 6.90, p-value < 0.0001 for the help-supporting regions). As to each age group, avatars in the commercial transaction region were significantly more likely to perform high-active actions than avatars in the help-supporting regions (t = 3.03, p-value = 0.0013 for young-aged avatars; t = 10.15, p-value <0.0001 for old-aged avatars), while avatars in the help-supporting regions were significantly more likely to perform low-active actions than avatars in the commercial transaction region (t = 1.69, p-value = 0.0462 for young-aged avatars; t = 6.26, p-value < 0.0001 for old-aged avatars). Therefore, H6 was supported in all cases.

Tables 4.7 and 4.8 show the three-factor interaction effect testing results for high-active and low-active actions respectively.

Table 4.7 Three-factor interaction effect testing results for high-active actions (H7)

High-Activity		N	Mean	Std. Dev.	F	p-value
*Virtual Gender * Virtual Age * Region Theme*					24.35	<0.0001
M + Y + C (MYC)	(MYC vs. FYC: t = 1.75, p-value = 0.0415)	55	0.60	0.28		
F + Y + C (FYC)	(MOC vs. FOC : t = 2.00, p-value = 0.0235)	99	0.51	0.30		
M + O + C (MOC)	(MYH vs. FYH : t = 1.60, p-value = 0.0552)*	148	0.52	0.29		
F + O + C (FOC)	(MOH vs. FOH: t = 3.32, p-value = 0.0005)	293	0.46	0.27		
M + Y + H (MYH)	(MYC vs. MOC: t = 1.73, p-value = 0.0433)	163	0.49	0.23		
F + Y + H (FYH)	(FYC vs. FOC: t = 1.46, p-value = 0.0732)*	158	0.45	0.23		
M + O + H (MOH)	(MYH vs. MOH: t = 5.51, p-value < 0.0001)	171	0.36	0.21		
F + O + H (FOH)	(FYH vs. FOH: t = 7.51, p-value = < 0.0001)	247	0.29	0.18		
	(MYC vs. MYH: t = 2.85, p-value = 0.0024)					
	(FYC vs. FYH: t = 2.00, p-value = 0.0235)					
	(MOC vs. MOH: t = 5.95, p-value < 0.0001)					
	(FOC vs. FOH: t = 8.64, p-value < 0.0001)					

Note. * Significant at 0.1 level. All the others are significant at 0.05 level.

Table 4.8 Three-factor interaction effect testing results for low-active actions (H7)

Low-Activity		N	Mean	Std. Dev.	F	p-value
*Virtual Gender * Virtual Age * Region Theme*					11.51	<0.0001
M + Y + C (MYC)	(MYC vs. FYC: t = 1.32, p-value = 0.0951)*	55	0.73	0.17		
F + Y + C (FYC)	(MOC vs. FOC : t = 1.47, p-value = 0.0714)*	99	0.76	0.18		
M + O + C (MOC)	(MYH vs. FYH : t = 0.55, p-value = 0.2927)**	148	0.77	0.16		
F + O + C (FOC)	(MOH vs. FOH: t = 2.12, p-value = 0.0175)	293	0.79	0.15		
M + Y + H (MYH)	(MYC vs. MOC: t = 1.73, p-value = 0.0436)	163	0.77	0.16		
F + Y + H (FYH)	(FYC vs. FOC: t = 1.63, p-value = 0.0521)*	158	0.78	0.14		
M + O + H (MOH)	(MYH vs. MOH: t = 3.77, p-value = 0.0001)	171	0.83	0.13		
F + O + H (FOH)	(FYH vs. FOH: t = 5.58, p-value < 0.0001)	247	0.86	0.13		
	(MYC vs. MYH: t = 1.80, p-value = 0.0376)					
	(FYC vs. FYH: t = 0.85, p-value = 0.1975)**					
	(MOC vs. MOH: t = 3.71, p-value = 0.0001)					
	(FOC vs. FOH: t = 5.28, p-value < 0.0001)					

Note. ** Not significant. * Significant at 0.1 level. All the others are significant at 0.05 level.

For both high-active and low-active actions, the three-factor interaction effect (F = 24.35, p-value < 0.0001 and F = 11.55, p-value < 0.0001) was significant. When comparing virtual gender differences, among all four cases (two levels of virtual age * two levels of region theme), male avatars were more likely to perform high-active actions than female avatars in three cases at 0.05 significance level, and female avatars were more likely to perform low-active actions than female avatars in one case (old-aged avatars in help-supporting regions) at 0.05 significance level.

When comparing virtual age differences, among all four cases (two levels of virtual gender * two levels of region theme), young-aged avatars were more likely to perform high-active actions than old-aged avatars in three cases at 0.05 significance level and barely significant in the other case (female avatars in the commercial transaction region, at 0.1 significance level); old-aged avatars were more likely to perform low-active actions than young-aged avatars in the same three cases at 0.05 significance level and barely significant in the group of female avatars in the commercial transaction region (at 0.1 significance level).

When comparing region theme differences, avatars in the commercial transaction region were significantly more likely to perform high-active actions than avatars in the help-supporting regions in all four cases (two levels of virtual gender * two levels of virtual age). In three cases (except for young-aged female avatars), avatars in the help-supporting regions were significantly more likely to perform low-active actions than avatars in the commercial transaction region. Therefore, H7 was supported in most cases.

4.7 Conclusions and Future Directions

Virtual worlds have rich social media data with great potential and importance to researchers interested in various areas, such as business, social science, communication, and education. In order to fully use the benefits of virtual worlds, it is important to know whether people’s behavior is consistent across both the virtual world and the physical world. However, due to some technical difficulties of this new media, especially data collection, little research has been done toward addressing this overarching research question. As exploratory research, this study examined whether real-life social norms hold in the virtual world and what the major factors are that can influence people's behavior in the virtual world.

Guided by the theories of social presence, social role, and gender role, the main effects and interaction effects of avatar virtual gender, virtual age, and region theme on avatars' physical activities were examined. To enable the analyses, an integrated technical framework for avatar data collection was developed by combining an improved bot-based approach and a spider-based approach.

The experimental results indicated that, in general, male avatars were more (less) likely to perform high-active (low-active) actions than female avatars; young-aged avatars were more (less) likely to perform high-active (low-active) actions than old-aged avatars; avatars in commercial transaction regions were more (less) likely to perform high-active (low-active) actions than avatars in non-profit, help-supporting regions. The following sections discussed the research contributions, implications, and future directions of this study.

4.7.1 Research Contributions

This study has significant research contributions to both the technical and behavioral research fields of IS.

For technical research, this study developed an integrated framework for data collection in the virtual world. Currently, for researchers who are interested in virtual worlds, data collection is a critical problem. It is always difficult to collect avatar related data since most virtual worlds do not provide easy data collection function. The integrated data collection framework developed in this study provides a systematic way for collecting various types of avatar behavior and profile data. In addition, the improved bot-based approach in the framework has addressed the limitations of the existing bot-based approach developed by Yee et al. (2007). As summarized in Table 4.1, by utilizing the LibOpenMetaverse, the improved bot-based approach allows each bot to conduct a much larger-scale data collection in a much faster manner. Further, by combining the improved bot-based approach with a spider-based approach, the framework is able to collect not only avatar behavioral data but also avatar profile data, which hadn't been done in previous research.

For behavioral research, this study examined and identified similar gender and age differences toward activities in the virtual world compared to the physical world. Similarly as in the physical world, in general, male avatars were more (less) likely to perform high-active (low-active) actions than female avatars, while young-aged avatars were more (less) likely to perform high-active (low-active) actions than old-aged avatars. In addition, this study introduced region theme as an important context factor for virtual world, and showed that it also had a significant impact on avatar activity. Specially, avatars in commercial transaction regions were more (less) likely to perform high-active (low-active) actions than avatars in non-profit, help-supporting regions. Furthermore, the study examined both the main and interaction effects of virtual gender, virtual age, and region theme on avatars' physical activities, which had not been examined before.

4.7.2 Implications

The results of the present study also have implications for future research on virtual worlds. First, the study provided preliminary empirical evidence toward answering the overarching research question of whether people generally demonstrate consistent behavior across the virtual world and the physical world from a particular perspective of avatars' physical activities. The supportive analysis results encourage future studies that cannot (or not easily) be conducted in real-life settings to be conducted in virtual worlds. Similar results or patterns could be expected compared with using real human subjects.

Second, it suggests that when conducting experiments using avatars as research subjects, it is better to consider and control the factors of avatar virtual gender, virtual age, and region theme. These factors could moderate avatar behavior.

In addition, different strategies need to be designed in order to attract different targeted avatar groups based on their characteristics. For example, since old-aged (more experienced) avatars are less physically active in general, they are more likely to stay in a certain location without moving around frequently. Therefore, for virtual companies or stores that focus on providing products or services to old-aged avatars, it is better to send out sales clerks or put up posters in multiple places instead of in only the location of the company or store itself. In this way, with enhanced user awareness, it will be easier to attract potential customers.

4.7.3 Future Research Directions

This study has a few limitations that can be addressed in future research. First, the gender information used in this study is the virtual gender of avatars instead of the real gender of people who control those avatars. The two types of gender may not be the same. For example, a man in the physical world could use a female avatar to represent him in the virtual world. Given that male avatars behaved like males and female avatars behaved liked females, it is unlikely that crossing genders had an impact on the results. However, to truly examine and compare gender differences in the virtual world vs. the physical world, it is important to link an avatar's virtual gender to the real gender of the person whom the avatar represents. Thus, future research needs to distinguish avatars that have gender information that is both consistent with and different from that of the real persons they represent, and identify the characteristics and behavior differences of each group.

Another limitation is the measure of avatar virtual ages which are based on how long avatars have existed in the virtual world. It actually measures the experience of avatars instead of the real ages of people who control those avatars. For example, an adult may have recently registered in the virtual world, thus having a young-aged avatar; while a child could have been registered in the virtual world for a long time, thus having an old-aged avatar. Future research needs to link an avatar's virtual age to the real age of the person whom the avatar represents to further investigate and compare age impacts, in contrast to experience impacts, on behavior in the virtual and physical worlds.

In addition, as a rich communication media, virtual worlds provide various communication channels, including verbal or nonverbal. The present study focused on examining avatar activity differences based on nonverbal data. Future research could examine how to collect verbal cues from the virtual world and how to leverage both verbal and non-verbal cues together to study avatar behavior. In addition to avatars' physical activities, future research can also investigate the impacts of avatar gender, age, and region theme on other types of avatar behavior, such as sentiments, emotions, and personality (positivity, friendliness, altruism, etc.).

Besides the three factors examined in this study, future research can explore other factors that may also influence avatar behavior. Besides analyzing individual avatar's activity, future research can also examine how to build the social interaction networks among avatars, and compare their unique characteristics (if any) to real-life social networks.

CHAPTER 5. AVATAR GENDER AND AGE DIFFERENCES IN HELP-SEEKING INTERACTIONS IN THE VIRTUAL WORLD

5.1 Introduction

Chapter 4 analyzed the effects of avatar virtual gender, virtual age, and region theme on avatars' physical activities in the virtual world. Following one of the future research directions suggested in Chapter 4, this Chapter examines the social interaction networks among avatars. Specifically, avatars' gender and age differences in their social interactions in help-seeking regions in the virtual world are studied. Both the overall gender and age difference analyses and detailed investigations by comparing three types of interaction networks based on gender or age are conducted.

Virtual worlds have been growing to attract more and more attention. It is expected that virtual worlds will grow in societal importance and will influence various aspects of people's lives such as education, collaboration, business and research (Messinger et al. 2009). In addition, a virtual world serves as a new place to do business effectively in new ways and has been bringing a lot of business opportunities. For example, the economy of the most popular virtual world, Second Life, totals $567 million US dollars in 2009 with 65% growth over 2008. Xstreet SL, the web marketplace for Second Life virtual goods, had gross sales of L$485 million, or approximately US$1.9 million in the last quarter of 2009 (http://blogs.secondlife.com/). The emerging virtual world market could reach billions of dollars in the coming years.

Researchers are facing various types of issues related to this new type of world, including business, social, political, communication, educational, technical, ethical, legal issues, etc (Gordon et al. 2009; Hendaoui et al. 2008). It is important to study and understand people's behavior in this new media. According to gender role theory, in the physical world, people with different genders and/or ages always behave differently. Both gender and age are treated as having important moderating effects on various types of human behavior. However, few studies have examined their impacts on avatar behavior especially their social interactions in virtual worlds.

Different from the physical world, as a complex platform with technical supports, virtual worlds often impose a considerable learning curve on users, especially for creating virtual items (Diversified Media Design et al. 2007; Messinger et al. 2009). Therefore, seeking help is one of the most important avatar behaviors in virtual worlds. In this study, I examined avatars' gender and age differences in their social interactions in help-seeking regions in the virtual world. Both the overall gender and age difference analyses and detailed investigations by comparing three types of interaction networks based on gender or age were conducted.

The remainder of this chapter is organized as follows. Section 5.2 provides the related literature and theoretical background. Section 5.3 lists the research questions of this study. Section 5.4 provides the research hypotheses. Section 5.5 describes the research method. Sections 5.6 and 5.7 show the detailed data analyses and results. Section 5.8 highlights the research contributions and discussions of future research directions.

5.2 Related Literature and Theories

5.2.1 Gender Roles

Gender roles are considered as a set of social and behavioral norms appropriate for individuals of a specific gender. Gender role theory (Eagly and Karau 1991) states that, in general, gender differences in behavior and personality are socially constructed. Gender role theory was derived from role theory (Biddle 1986) that posits people act according to expectations from themselves and others.

Gender role theory suggests that “individuals internalize cultural expectations about their gender because social pressures external to the individual favor behavior consistent with their prescribed gender role” (Kidder 2002, p. 630). People use these expectations to categorize themselves and others, and thus these factors guide individuals toward their own behavior (Malcom 2003; Palapattu et al. 2006).

In addition, gender roles have been found to be moderated by age (Malcom 2003). As two major markers of social identity, both gender and age are important factors to consider when examining people’s behavior differences (Blascovich et al. 2002).

5.2.2 Gender and Age Differences in Social Interactions

Previous research on gender differences in social interactions found that, in general, women had larger networks with more friends than men (Landman-Peeters et al. 2005; Shye et al. 1995; Tonge 2010), and provided and received more support from members of their networks than men (Antonucci and Akiyama 1987; Landman-Peeters et al. 2005; Shye et al. 1995).

Most research has consistently found that women's social interactions mainly contained supportive behaviors that were influenced by feminine social role expectations, while men's social interactions were mainly based on expectations of independence (Antonucci and Akiyama 1987; Bowling 1991; Shye et al. 1995). These gender differences in social interactions have been shown to be generally consistent across different age groups (Sears et al. 2009; Shye et al. 1995).

5.2.3 Gender and Age Differences in Help-Seeking Scenario

Previous research found significant gender differences in people's help-seeking behavior. In most cases, women were more willing to seek professional and academic help than men (Solberg et al. 1994; Yeh 2002). Men were less likely to seek help than women. Even when men needed help, they were more willing to seek help from their female friends than male friends (Sears et al. 2009).

In both formal (e.g., from professionals such as teachers, doctors, social workers, and psychologists) and informal (e.g., from friends) help channels, consistent results were found that women had significantly more willingness to seek help than men (Grinstein-Weiss et al. 2005).

Such gender differences can be attributed to the different role expectations for men and women (Sears et al. 2009). Men's social role expectations value achievement and independence, while women's social role expectations encourage collaboration, dependence, and emotional expression (Grinstein-Weiss et al. 2005; Sears et al. 2009). Consequently, compared with women, men were more often treated as autonomous problem solvers, thus becoming reluctant to seek help (Grinstein-Weiss et al. 2005). Previous research also found that help-seeking behavior tended to increase with an increase in age (Grinstein-Weiss et al. 2005; Yeh 2002).

5.2.4 Previous Research on Gender and Age Differences in Virtual Worlds

For gender differences in virtual worlds, previous research found that the two genders had social space differences. Male avatars tended to keep larger interpersonal distance with other male avatars compared to the interpersonal distance between two female avatars (Yee et al. 2007). They also found that male avatars tended to have less eye contact with other male avatars when compared to the frequency of eye contact between two female avatars.

For virtual age differences, Harris et al. (2009) found that in general avatars' high-active activities declined with increased usage experience (as they grew older). One limitation of their study is that they conducted a lab experiment with 80 students with motivation to get course credit instead of Second Life residents.

The results of the study in Chapter 4 on analyzing avatar physical activity differences found that male avatars tended to perform more active behaviors than female avatars, and young-aged avatars tended to perform more active behaviors than old-aged avatars.

5.2.5 Previous Research on Avatar Social Interactions in Virtual Worlds

Previous research examined how avatars' relationships were affected by different environmental factors in massively multiplayer online game such as EverQuest II based on analyzing user logs (Shim et al. 2011). For example, Keegan et al. (2010) found significant differences in the trade networks of gold famers and non-gold farmers in EverQuest II by measuring social network metrics such as in-degree, out-degree, and centrality. They found that gold famers had significantly lower in-degrees and out-degrees compared with non-gold famers, indicating that gold famers typically tended to conduct transactions with only a few trusted characters. Kawale et al. (2009) studied the problem of player churn in EverQuest II by analyzing player engagement and network statistics such as the number of neighbors and number of churner neighbors.

For social network analysis in Second Life, previous research built "human-to-human" and "human-to-bot" interaction networks and found significant differences in social network metrics such as degree, clustering coefficient, and betweenness centrality (Varvello and Voelker 2010). The interaction networks were modeled as undirected graphs. Two avatars were assumed to be interacted with each other when their Euclidean distance was less than 5 meters in a region. They suggested that nodes with much higher degrees and betweenness centrality scores compared with others most likely correspond to bots.

5.3 Research Questions

No existing study has examined gender and age differences of social interactions for help-seeking in virtual worlds. This study seeks to address the following research questions:

1. How to examine social interactions between gender (age) groups?
2. What's the social interaction difference between male and female avatars when seeking help?
3. What's the social interaction difference between young-aged and old-aged avatars when seeking help?

5.4 Research Hypotheses

According to gender role theory, males and females are expected to have different behaviors because the different role expectations for the two genders. Previous research examined gender and age differences in people's physical world social interactions and consistently found that women had more friends with more supportive behaviors in their social interactions than men across all age groups (Landman-Peeters et al. 2005; Shye et al. 1995; Tonge 2010) that resulted from their social role expectations of being supportive compared to men's social role expectations of being independent (Antonucci and Akiyama 1987; Bowling 1991; Shye et al. 1995). Similar patterns could be expected in the virtual world - female avatars might have more social interactions than male avatars.

In terms of help-seeking behavior, previous research also found significant gender and age differences in the physical world. In most cases, women have been found to be more willing than men to seek help from others (Solberg et al. 1994; Yeh 2002). It is also reported that help-seeking behavior tended to increase with age (Grinstein-Weiss et al. 2005; Yeh 2002). Therefore, it could be expected that gender and age are important factors to influence avatars' help seeking behavioral in the virtual world. Therefore, I posit that male and female avatars are different in help seeking interactions. Similarly, young-aged and old-aged avatars are also expected to be different in help seeking interactions.

> H1. Male and female avatars are different in help seeking interactions in the virtual world.

> H2: Young-aged and old-aged avatars are different in help seeking interactions in the virtual world.

To compare avatar social interactions, previous research has developed different types of interaction networks, such as the gold famers vs. non-gold farmers networks (Keegan et al. 2010), and "human-human" vs. "human-bots" networks (Varvello and Voelker 2010). Significant differences were found among different types of networks. In terms of gender, the network based on interactions among the same gender (either male or female) is expected to be different from the network based on interactions between the two different genders. Similarly, in terms of age, the network based on interactions among the same age group (either young-aged avatars or old-aged avatars) is different from the network based on interactions between the two different age groups. Therefore, I hypothesize:

> H3. Help seeking networks containing interactions only among male avatars, only among female avatars, and between male and female avatars are different.

H4: Help seeking networks containing interactions only among young-aged avatars, only among old-aged avatars, and between young-aged and old-aged avatars are different.

5.5 Research Method

5.5.1 Data Collection

Data collection was conducted using the integrated framework developed in Chapter 4. The integrated framework contains two parts: improved bot-based approach and spider-based approach (see Figure 4.2 in Chapter 4).

A bot was created as a real avatar, and logged in Second Life using LibOpenMetaverse (http://openmetaverse.org/) as client to automatically collect behavioral data of all other avatars in a given region. Two most popular help-seeking regions with largest numbers of participants were used as the study sites. They are “Help Island Public” and “Moose Beach.” One bot was sent to each region (65536 m^2, 16.1943 acres). Five-second sampling resolution was used to collect avatar behavior records (Harris et al. 2009; Yee and Bailenson 2008). A record includes avatar ID, avatar name, avatar’s group in Second Life, region name, avatar position (x, y, and z coordinates), avatar face angle, action ID, action name, and time stamp.

Data was collected from the region “Help Island Public” from 06/21/2010 to 07/07/2010. A total of 1,379,964 records from 4,558 unique avatars were collected. Data was collected from the region “Moose Beach” from 06/18/2010 to 07/08/2010. A total of 1,474,192 records from 6,110 unique avatars were collected.

The spider-based approach was then used to collect avatar profile data. The profile page of each avatar was collected by matching the avatar name collected in the previous step. Avatar birthday information was extracted from the profile pages. Avatars' virtual ages were calculated accordingly. Virtual gender can be inferred from the data fields of avatar name, self introduction, and profile picture listed on the profile pages. Two coders did the avatar gender coding based on these data fields. Only the avatars whose gender information, either male or female, was able to be clearly decided without any disagreement or confusion among both coders were kept.

5.5.2 Measures

Two avatars are assumed to be interacting when their Euclidean distance is less than 5 meters in a region. I followed this assumption adopted in previous research on avatar social interaction analysis in Second Life (Varvello and Voelker 2010). The 5-meter distance threshold was suggested by previous literature (Varvello and Voelker 2010), and is also consistent with my own observation. The Euclidean distance D is calculated as: $D = \sqrt[2]{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$, where ($x_1$, y_1, z_1) and (x_2, y_2, z_2) are the positions of two avatars in a given region.

To measure avatar social interactions, the most widely used social network related metrics are adopted, including degree, betweenness centrality, Hyperlink-Induced Topic Search (HITS), and PageRank.

Degree: The degree of a node $v \in V(G)$ is the number of edges linked to it. In an undirected graph, in-degree and out-degree are equal. It represents the number of unique avatars interacted with avatar *v*.

Betweenness centrality: It measures the extent to which a node lies on the shortest paths between every other node (Freeman 1977; Freeman 1979). Nodes around the periphery of a network would typically have a low betweenness centrality. A high betweenness centrality might suggest that the node is connecting various parts of a network together.

Hyperlink-Induced Topic Search (HITS): HITS (Kleinberg 1999) is a link analysis algorithm. Each element in a linked structure is ranked according to its degree of "authority" and "hub." Elements with a large number of incoming links are considered as "authorities" while elements with a large number of outgoing links are considered as "hubs." The authority score is calculated as $HITS_{Authority}(s) = \sum_{t \in In(s)} HITS_{Hub}(t)$, and the hub score is calculated as $HITS_{Hub}(s) = \sum_{t \in Out(s)} HITS_{Authority}(t)$.

PageRank: PageRank (Brin and Page 1998) is a link analysis algorithm aims to measure the relative importance of each element (e.g., a Web page) in a set of linked elements (e.g., the World Wide Web). The PageRank score is calculated by integrating the impacts of both incoming and outgoing links. It is calculated as $PR(s) = (1 - d) + d * \sum_{t \in In(s)} \frac{PR(t)}{|Out(t)|}$, where *d* is a parameter with value between 0 and 1.

Both HITS and PageRank were originally developed for directed networks (graphs). But they can also be applied to measure undirected graphs by treating the in-degree of a node as equal to the out-degree of the node (both are treated as the "degree" of the node) (Mihalcea 2004; White and Smyth 2003). Thus, for undirected graphs, the HITS authority score and hub score of a node are the same (White and Smyth 2003).

In this study, all the above measures were calculated using Java Universal Network/Graph Framework (JUNG; http://jung.sourceforge.net/), an open source graph modeling and visualization package in Java (O'Madadhain et al. 2005).

5.6 Avatar Gender Difference Analysis and Results

To test H1 and H3, both of which are related to the avatar gender difference, the overall (H1) and detailed analyses (H3) were conducted respectively. A total of 2,090 avatars with clear gender information identified by the coders were obtained from the region Help Island Public. Among them, 995 were male avatars and 1,095 were female avatars. A total of 2,967 avatars with clear gender information identified by the coders were obtained from the region Moose Beach. Among them, 1,347 were male avatars and 1,620 were female avatars.

5.6.1 Overall Avatar Gender Difference Analysis and Results

To test H1, I first conducted ANOVA analysis to examine the overall impact of gender in avatar social interactions in both regions. Tables 5.1 and 5.2 summarize the testing results.

Table 5.1 ANOVA test result of the overall gender difference in the region Help Island Public (H1)

	Male (N=995)		Female(N=1,095)		F	p-value
	Mean	Std. dev	Mean	Std. dev		
Degree	13.481	22.379	14.426	24.696	0.83	0.362
Betweeness Centrality	$1.717*10^{-3}$	$7.944*10^{-3}$	$2.008*10^{-3}$	$9.276*10^{-3}$	0.59	0.444
HITS authority/hub score	$1.101*10^{-2}$	$1.703*10^{-2}$	$1.147*10^{-2}$	$1.912*10^{-2}$	0.33	0.567
PageRank score	$4.459*10^{-4}$	$7.335*10^{-4}$	$4.762*10^{-4}$	$8.097*10^{-4}$	0.80	0.372

Table 5.2 ANOVA test result of the overall gender difference in the region Moose Beach (H1)

	Male (N=1,347)		Female(N=1,620)		F	p-value
	Mean	Std. dev	Mean	Std. dev		
Degree	16.039	30.517	15.394	25.413	0.39	0.530
Betweeness Centrality	$1.549*10^{-3}$	$8.612*10^{-3}$	$1.265*10^{-3}$	$5.252*10^{-3}$	1.22	0.270
HITS authority/hub score	$9.839*10^{-3}$	$1.659*10^{-2}$	$9.146*10^{-3}$	$1.486*10^{-2}$	1.44	0.230
PageRank score	$3.408*10^{-4}$	$6.455*10^{-4}$	$3.265*10^{-4}$	$5.376*10^{-4}$	0.43	0.512

The ANOVA test results did not show significant differences between male and female avatars in terms of degree, betweenness centrality, HITS, or PageRank scores in both regions. This indicated that there was no significant difference between male avatars' interactions with other avatars (including both male and female avatars) and female avatars' interactions with other avatars (including both male and female avatars) in general. Thus, H1 was not supported.

5.6.2 Detailed Avatar Gender Difference Analysis and Results

To test the detailed avatar gender difference (H3), three distinct types of avatar interaction networks were conducted, including male-male interaction network, female-female interaction network, and male-female/female-male interaction network.

- Male-male interaction network contains all the interactions among male avatars. Each node in the network is a unique male avatar.
- Female-female interaction network contains all the interactions among female avatars. Each node in the network is a unique female avatar.
- Male-female/female-male interaction network contains all the interactions between a male avatar and a female avatar. For the two nodes linked by each edge, one represents a male avatar and the other represents a female avatar.

Table 5.3 lists the descriptive statistics of the three interaction networks from the region Help Island Public. Bonferroni adjusted t-tests were conducted to compare the network characteristics among these three interaction networks (see Table 5.4).

Table 5.3 Descriptive statistics of male-male, female-female, and male-female/female-male interaction networks from the region Help Island Public

	Male-Male Network (N=880)		Female-Female Network (N=1,016)		Male-Female/Female-Male Network (N=1,909)	
	Mean	**Std. dev**	**Mean**	**Std. dev**	**Mean**	**Std. dev**
Degree	6.691	10.601	8.006	13.018	7.414	11.843
Betweeness Centrality	$5.304*10^{-3}$	$2.153*10^{-2}$	$4.543*10^{-3}$	$1.753*10^{-2}$	$2.950*10^{-3}$	$1.139*10^{-2}$
HITS authority/hub score	$1.842*10^{-2}$	$2.825*10^{-2}$	$1.616*10^{-2}$	$2.690*10^{-2}$	$1.223*10^{-2}$	$1.935*10^{-2}$
PageRank score	$1.136*10^{-3}$	$1.763*10^{-3}$	$9.843*10^{-4}$	$1.591*10^{-3}$	$5.238*10^{-4}$	$8.304*10^{-4}$

Table 5.4 Bonferroni adjusted t-test results of male-male, female-female, and male-female/female-male interaction networks from the region Help Island Public (H3)

Measure	Comparison	t-stat	p-value
Degree	Male-Male < Female-Female	2.42	0.0233*
	Male-Male < Male-Female/Female-Male	1.61	0.1604
	Female-Female > Male-Female/Female-Male	1.21	0.3416
Betweeness Centrality	Male-Male > Female-Female	0.84	0.6050
	Male-Male > Male-Female/Female-Male	3.05	0.0035**
	Female-Female > Male-Female/Female-Male	2.62	0.0135*
HITS authority/hub score	Male-Male > Female-Female	1.79	0.1115
	Male-Male > Male-Female/Female-Male	5.90	<.0001**
	Female-Female > Male-Female/Female-Male	4.12	<.0001**
PageRank score	Male-Male > Female-Female	1.96	0.0752
	Male-Male > Male-Female/Female-Male	9.82	<.0001**
	Female-Female > Male-Female/Female-Male	8.62	<.0001**

Note. p-values are Bonferroni corrected p-values. *Significant at 0.05 level. ** Significant at 0.01 level.

As shown in Table 5.4, the degree of female-female interaction network was significantly larger than that of the male-male interaction network ($t=2.42$, $p=0.0233$). This indicated that compared with the interactions among male avatars (with average degree of 6.691), on average female avatars had more interactions (with average degree of 8.006) with other female avatars. There was no significant difference between the male-male (female-female) interaction network and male-female/female-male interaction network.

The betweenness centrality scores of both male-male ($t=3.05$, $p=0.0035$) and female-female ($t=2.62$, $p=0.0135$) interaction networks were significantly higher than that of the male-female/female-male interaction network. This indicated that the interaction networks of the same gender tended to be more centralized than the network of interactions between the two different genders. This also suggested that there might be some male (female) avatars with whom a lot of other avatars in the same gender often interacted. There was no significant difference between the male-male and female-female interaction networks.

Similar patterns were observed for the HITS authority/hub score and the PageRank score. Both male-male ($t=5.90$, $p<0.0001$) and female-female ($t=4.12$, $p<0.0001$) interaction networks had significantly higher HITS authority/hub score than the male-female/female-male interaction network. Both male-male ($t=9.82$, $p<0.0001$) and female-female ($t=8.62$, $p<0.0001$) interaction networks had significantly higher PageRank score than the male-female/female-male interaction network. The results indicated that on average the interaction networks of the same gender tended to be more centralized than the interaction network of the two different genders. This also suggested that there could be some important male (female) avatars with whom a lot of other avatars in the same gender often interacted, while in the male-female/female-male interaction network, such important authorities/hubs could be fewer.

Table 5.5 lists the descriptive statistics about the region Moose Beach. Table 5.6 shows the Bonferroni adjusted t-test results for the region Moose Beach.

Table 5.5 Descriptive statistics of male-male, female-female, and male-female/female-male interaction networks from the region Moose Beach

	Male-Male Network (N=1,221)		**Female-Female Network (N=1,513)**		**Male-Female/Female-Male Network (N=2,762)**	
	Mean	**Std. dev**	**Mean**	**Std. dev**	**Mean**	**Std. dev**
Degree	8.082	14.704	8.662	13.504	8.329	14.485
Betweeness Centrality	$3.651*10^{-3}$	$1.750*10^{-2}$	$3.192*10^{-3}$	$1.159*10^{-2}$	$2.038*10^{-3}$	$8.535*10^{-3}$
HITS authority/hub score	$1.526*10^{-2}$	$2.422*10^{-2}$	$1.319*10^{-2}$	$2.208*10^{-2}$	$9.991*10^{-3}$	$1.620*10^{-2}$
PageRank score	$8.190*10^{-4}$	$1.476*10^{-3}$	$6.609*10^{-4}$	$1.019*10^{-3}$	$3.621*10^{-4}$	$6.156*10^{-4}$

Table 5.6 Bonferroni adjusted t-test results of male-male, female-female, and male-female/female-male interaction networks from the region Moose Beach (H3)

Measure	Comparison	t-stat	p-value
Degree	Male-Male < Female-Female	1.06	0.4313
	Male-Male < Male-Female/Female-Male	0.49	0.9324
	Female-Female > Male-Female/Female-Male	0.75	0.6779
Betweeness Centrality	Male-Male > Female-Female	0.79	0.6459
	Male-Male > Male-Female/Female-Male	3.06	0.0033**
	Female-Female > Male-Female/Female-Male	3.40	0.0011**
HITS authority/hub score	Male-Male > Female-Female	2.31	0.0315*
	Male-Male > Male-Female/Female-Male	6.94	<.0001**
	Female-Female > Male-Female/Female-Male	4.95	<.0001**
PageRank score	Male-Male > Female-Female	3.18	0.0023**
	Male-Male > Male-Female/Female-Male	10.43	<.0001**
	Female-Female > Male-Female/Female-Male	10.42	<.0001**

Note. p-values are Bonferroni corrected p-values. *Significant at 0.05 level. ** Significant at 0.01 level.

In the region Moose Beach, the differences of the average degrees among the three types of networks were not significant.

Testing the result of betweenness centrality in this region was consistent with that in the region Help Island Public. The betweenness centrality scores of both male-male (t=3.06, p=0.0033) and female-female (t=3.40, p=0.0011) interaction networks were significantly higher than that of the male-female/female-male interaction network. However, there was no significant difference between the male-male and female-female interaction networks.

Testing the results of the HITS authority/hub score and PageRank score were similar as those obtained in the region Help Island Public. Both male-male ($t=6.94$, $p<0.0001$) and female-female ($t=4.95$, $p<0.0001$) interaction networks had significantly higher HITS authority/hub score than the male-female/female-male interaction network. Both male-male ($t=10.43$, $p<0.0001$) and female-female ($t=10.42$, $p<0.0001$) interaction networks had significantly higher PageRank score than the male-female/female-male interaction network. In addition, significant difference between the male-male and female-female interaction networks was also obtained in this region. The male-male interaction network had significantly higher HITS authority/hub score ($t=2.31$, $p=0.0315$) as well as PageRank score ($t=3.18$, $p=0.0023$) than the female-female interaction network.

Overall, H3 was supported in most cases. The degree of female-female interaction network was significantly larger than that of the male-male interaction network in one region in the study. Both male-male and female-female interaction networks had significantly higher betweenness centrality, HITS, and PageRank scores than the male-female/female-male interaction network, indicating that the interaction networks of the same gender tended to be more centralized than the network of interactions between the two genders.

5.7 Avatar Age Difference Analysis and Results

This section presents detailed descriptions and discussions of the testing of H2 and H4. Three months was used as the cutoff between young-aged avatars and old-aged avatars (Zhang et al. 2010). A total of 4,062 avatars with clear age information were obtained from the region Help Island Public. Among them, there were 2,391 young-aged avatars and 1,671 old-aged avatars. A total of 5,516 avatars with clear age information were obtained from the region Moose Beach. Among them, there were 2,754 young-aged avatars and 2,762 old-aged avatars.

5.7.1 Overall Avatar Age Difference Analysis and Results

To test H2, ANOVA tests were conducted to examine the overall impact of age in avatar social interactions in both regions. Tables 5.7 and 5.8 summarize the testing results.

Table 5.7 ANOVA test result of the overall age difference in the region Help Island Public (H2)

	Young (N=2,391)		**Old (N=1,671)**		**F**	**p-value**
	Mean	**Std. dev**	**Mean**	**Std. dev**		
Degree	17.272	23.382	24.007	44.218	39.62	<.0001**
Betweeness Centrality	$6.557*10^{-4}$	$3.322*10^{-3}$	$1.620*10^{-3}$	$8.152*10^{-3}$	27.05	<.0001**
HITS authority/hub score	$7.212*10^{-3}$	$9.418*10^{-3}$	$9.513*10^{-3}$	$1.682*10^{-2}$	30.88	<.0001**
PageRank score	$2.021*10^{-4}$	$2.7294*10^{-4}$	$2.806*10^{-4}$	$5.163*10^{-4}$	39.53	<.0001**

Note. *Significant at 0.05 level. ** Significant at 0.01 level.

Table 5.8 ANOVA test result of the overall age difference in the region Moose Beach (H2)

	Young (N=2,754)		Old (N=2,762)		F	p-value
	Mean	**Std. dev**	**Mean**	**Std. dev**		
Degree	19.137	29.093	24.691	46.439	28.32	<.0001**
Betweeness Centrality	$4.981*10^{-4}$	$3.032*10^{-3}$	$1.027*10^{-3}$	$5.801*10^{-3}$	18.02	<.0001**
HITS authority/hub score	$6.191*10^{-3}$	$8.583*10^{-3}$	$7.354*10^{-3}$	$1.306*10^{-2}$	15.28	<.0001**
PageRank score	$1.482*10^{-4}$	$2.253*10^{-4}$	$1.913*10^{-4}$	$3.596*10^{-4}$	28.40	<.0001**

Note. *Significant at 0.05 level. ** Significant at 0.01 level.

For both regions, the ANOVA test results showed significantly differences between the two age groups in terms of degree, betweenness centrality, HITS authority/hub score, and RageRank score (all p-values < 0.0001). Thus, H2 was supported in all measurement dimensions.

On average, old-aged avatars had larger degrees than young-aged avatars, indicating that compared with young-aged avatars, old-aged avatars in general interacted with more avatars (no matter young or old). This may suggest that when avatars (no matter young or old) had problems, they were more likely to interact with old-aged avatars to seek help.

Old-aged avatars also had higher betweenness centrality, HITS authority/hub, and RageRank scores than the young-aged avatars. This means that old-aged avatars tended to be in the more important positions in the network. For example, they could be more likely to be in the center of the network and connect various parts of the network.

5.7.2 Detailed Avatar Age Difference Analysis and Results

Detailed avatar age difference analysis was conducted to test H4. Similar to the detailed avatar gender difference analysis, for age groups, three distinct types of avatar interaction networks were created and examined, including young-young interaction network, old-old interaction network, and young-old/old-young interaction network.

- Young-young interaction network contains all the interactions among young-aged avatars. Each node in the network is a unique young-aged avatar.
- Old-old interaction network contains all the interactions among old-aged avatars. Each node in the network is a unique old-aged avatar.
- Young-old/old-young interaction network contains all the interactions between a young-aged avatar and an old-aged avatar. For the two nodes linked by each edge, one represents a young-aged avatar and the other represents an old-aged avatar.

Table 5.9 lists the descriptive statistics of the three interaction networks from the region Help Island Public. Bonferroni adjusted t-tests were conducted to compare the network characteristics among these three interaction networks (see Table 5.10).

Table 5.9 Descriptive statistics of young-young, old-old, and young-old/old-young interaction networks from the region Help Island Public

	Young-Young Network (N=2,260)		Old-Old Network (N=1,579)		Young-Old/Old-Young Network (N=3,824)	
	Mean	**Std. dev**	**Mean**	**Std. dev**	**Mean**	**Std. dev**
Degree	8.603	11.984	11.598	19.731	10.379	17.390
Betweeness Centrality	$2.385*10^{-3}$	$1.227*10^{-2}$	$2.710*10^{-3}$	$1.089*10^{-2}$	$1.467*10^{-3}$	$6.457*10^{-3}$
HITS authority/hub score	$1.146*10^{-2}$	$1.764*10^{-2}$	$1.266*10^{-2}$	$2.176*10^{-2}$	$9.066*10^{-3}$	$1.339*10^{-2}$
PageRank score	$4.425*10^{-4}$	$6.108*10^{-4}$	$6.333*10^{-4}$	$1.071*10^{-3}$	$2.615*10^{-4}$	$3.998*10^{-4}$

Table 5.10 Bonferroni adjusted t-test results of young-young, old-old, and young-old/old-young interaction networks from the region Help Island Public (H4)

Measure	Comparison	t-stat	p-value
Degree	Young-Young < Old-Old	5.38	<.0001**
	Young-Young < Young-Old/Old-Young	4.70	<.0001**
	Old-Old > Young-Old/Old-Young	2.14	0.0489*
Betweeness Centrality	Young-Young < Old-Old	0.86	0.5823
	Young-Young > Young-Old/Old-Young	3.30	0.0015**
	Old-Old > Young-Old/Old-Young	4.24	<.0001**
HITS authority/hub score	Young-Young < Old-Old	1.82	0.1043
	Young-Young > Young-Old/Old-Young	5.57	<.0001**
	Old-Old > Young-Old/Old-Young	6.10	<.0001**
PageRank score	Young-Young < Old-Old	6.39	<.0001**
	Young-Young > Young-Old/Old-Young	12.58	<.0001**
	Old-Old > Young-Old/Old-Young	13.41	<.0001**

Note. p-values are Bonferroni corrected p-values. *Significant at 0.05 level. ** Significant at 0.01 level.

The old-old interaction network had the largest degree ($t=5.38$, $p<0.0001$ compared with young-young network; $t=2.14$, $p=0.0489$ compared with young-old/old-young network), followed by the young-old/old-young interaction network ($t=4.70$, $p<0.0001$ compared with young-young network). The young-young interaction network had the smallest degree. The result indicated that, in general, old-aged avatars had more interactions with old-aged avatars than with young-aged avatars. The same pattern existed for young-aged avatars – they also tended to have more interactions with old-aged avatars than with young-aged avatars. This may suggest that when seek help, both old-aged and young-aged avatars were more likely to interact with old-aged avatars.

The betweenness centrality scores of both young-young ($t=3.30$, $p=0.0015$) and old-old ($t=4.24$, $p=<0.0001$) interaction networks were significantly higher than that of the young-old/old-young interaction network. This indicated that the interaction networks of the same age group tended to be more centralized than the network of interactions between the two age groups. This may suggest that there might be some young-aged (old-aged) avatars with whom a lot of other avatars in the same age group often interacted. There was no significant difference between the young-young and old-old interaction networks.

Similar patterns were observed for the HITS authority/hub and PageRank scores. Both young-young ($t=5.57$, $p<0.0001$) and old-old ($t=6.10$, $p<0.0001$) interaction networks had significantly higher HITS authority/hub score than the young-old/old-young interaction network. Both young-young ($t=12.58$, $p<0.0001$) and old-old ($t=13.41$, p<0.0001) interaction networks had significantly higher PageRank score than the young-old/old-young interaction network. The results indicated that on average the interaction networks of the same age group tended to be more centralized than the interaction network of the two different age groups. This also suggested that there could be some important young-aged (old-aged) avatars with whom a lot of other avatars in the same age group often interacted, while in the young-old/old-young interaction network, such important authorities/hubs could be fewer.

Table 5.11 lists the descriptive statistics for Moose Beach region. Table 5.12 shows the Bonferroni adjusted t-test results for Moose Beach region.

Table 5.11 Descriptive statistics of young-young, old-old, and young-old/old-young interaction networks from the region Moose Beach

	Young-Young Network (N=2,588)		**Old-Old Network (N=2,662)**		**Young-Old/Old-Young Network (N=5,240)**	
	Mean	**Std. dev**	**Mean**	**Std. dev**	**Mean**	**Std. dev**
Degree	8.202	12.965	13.424	23.312	10.708	18.360
Betweeness Centrality	$2.076*10^{-3}$	$1.157*10^{-2}$	$1.649*10^{-3}$	$7.803*10^{-3}$	$1.094*10^{-3}$	$5.318*10^{-3}$
HITS authority/hub score	$1.047*10^{-2}$	$1.664*10^{-2}$	$9.728*10^{-3}$	$1.677*10^{-2}$	$7.375*10^{-3}$	$1.168*10^{-2}$
PageRank score	$3.864*10^{-4}$	$6.066*10^{-4}$	$3.757*10^{-4}$	$6.513*10^{-4}$	$1.908*10^{-4}$	$3.249*10^{-4}$

Table 5.12 Bonferroni adjusted t-test results of young-young, old-old, and young-old/old-young interaction networks from the region Moose Beach (H4)

Measure	Comparison	t-stat	p-value
Degree	Young-Young < Old-Old	10.07	<.0001**
	Young-Young < Young-Old/Old-Young	6.97	<.0001**
	Old-Old > Young-Old/Old-Young	5.24	<.0001**
Betweeness Centrality	Young-Young > Old-Old	1.56	0.1773
	Young-Young > Young-Old/Old-Young	4.11	<.0001**
	Old-Old > Young-Old/Old-Young	3.30	0.0015**
HITS authority/hub score	Young-Young > Old-Old	1.61	0.1617
	Young-Young > Young-Old/Old-Young	8.48	<.0001**
	Old-Old > Young-Old/Old-Young	6.48	<.0001**
PageRank score	Young-Young > Old-Old	0.62	0.8043
	Young-Young > Young-Old/Old-Young	15.35	<.0001**
	Old-Old > Young-Old/Old-Young	13.79	<.0001**

Note. p-values are Bonferroni corrected p-values. *Significant at 0.05 level. ** Significant at 0.01 level.

The old-old interaction network had the largest degree (t=10.07, p<0.0001 compared with young-young network; t=5.24, p=<0.0001 compared with young-old/old-young network), followed by the young-old/old-young interaction network (t=6.97, p<0.0001 compared with young-young network). This result is consistent with that obtained in the region Help Island Public, suggesting that when seeking help, both old-aged and young-aged avatars were more likely to interact with old-aged avatars.

Testing result of betweenness centrality in this region was consistent with that in the region Help Island Public. Although there was no significant difference between young-young and old-old interaction networks, the betweenness centrality scores of both young-young ($t=4.11$, $p=<0.0001$) and old-old ($t=3.30$, $p=0.0015$) interaction networks were significantly higher than that of the young-old/old-young interaction network.

The testing results of the HITS authority/hub and PageRank scores were similar as those obtained in the region Help Island Public. Although there was no significant difference between young-young and old-old interactions networks, both young-young ($t=8.48$, $p<0.0001$) and old-old ($t=6.48$, $p<0.0001$) interaction networks had significantly higher HITS authority/hub score than the young-old/old-young interaction network. Both young-young ($t=15.35$, $p<0.0001$) and old-old ($t=13.79$, $p<0.0001$) interaction networks had significantly higher PageRank score than the young-old/old-young interaction network.

Overall, H4 was supported in most cases. The old-old interaction network had the largest degree, followed by the young-old/old-young interaction network and the young-young interaction network. In addition, both young-young and old-old interaction networks had significantly higher betweenness centrality, HITS, and PageRank scores than the young-old/old-young interaction network. The results indicated that on average the interaction networks of the same age group tended to be more centralized than the interaction network of the two different age groups.

5.8 Research Contributions and Future Directions

5.8.1 Research Contributions

Although previous research has studied gender and age differences on people's help-seeking behavior in real-world settings, no study has examined such differences in virtual worlds. Previous virtual world studies examined gender differences in avatar interpersonal distances and eye contact patterns, and the impact of both gender and age in avatar physical activity differences. But no study has examined the gender and age differences in avatar social interactions. This study further tested gender role theory by examining avatar gender and age differences in their social interactions in help-seeking regions in the virtual world. To do this, both the overall gender and age difference analyses and detailed investigations were conducted. In the detailed analyses, three types of interaction networks based on gender or age were compared. The three types of networks based on gender are male-male, female-female, and male-female/female-male interaction networks. The three types of networks based on age are young-young, old-old, and young-old/old-young interaction networks.

For the overall gender and age difference analyses, the results showed that overall there was no significant difference between male and female avatars when seeking help in virtual world in terms of degree, betweenness centrality, HITS, or PageRank scores. But significant differences were obtained between young-aged and old-aged avatars in their help seeking interactions in virtual world in all measurement dimensions. Old-aged avatars had larger degree and higher betweenness centrality, HITS authority/hub, and RageRank scores than young-aged avatars. The results may indicate that when avatars had problems, they were more likely to interact with old-aged avatars to seek help, and old-aged avatars tended to be in the more important positions in the network.

For the detailed avatar gender difference analysis, the results showed that the degree of female-female interaction network was significantly larger than that of the male-male interaction network in one region in the study. This indicated that female avatars had more interactions with female avatars when seeking help compared with interactions among male avatars. In addition, both male-male and female-female interaction networks had significantly higher betweenness centrality, HITS, and PageRank scores than the male-female/female-male interaction network, indicating that the interaction networks of the same gender tended to be more centralized than the network of interactions between the two genders.

For the detailed avatar age difference analysis, the results showed that the old-old interaction network had the largest degree, followed by the young-old/old-young interaction network and the young-young interaction network. This suggests that when seeking help, both old-aged and young-aged avatars were more likely to interact with old-aged avatars. In addition, both young-young and old-old interaction networks had significantly higher betweenness centrality, HITS, and PageRank scores than the young-old/old-young interaction network. The results indicated that, on average, the interaction networks of the same age group tended to be more centralized than the interaction network of the two different age groups.

5.8.2 Implications and Future Research Directions

This study examined the impacts of avatar virtual gender and virtual age on avatar social interactions when seeking help. Future research can further explore avatar social interaction patterns in other areas such as marketing. Recent marketing literature has suggested examining the impacts of gender and other factors on consumer shopping behavior in channels (e.g., e-commerce websites) other than physical stores (Jackson et al. 2011). In the physical world, gender has been identified as an important factor influencing people’s shopping preferences, motives, and shopping value (Fischer and Arnold 1990; Jackson et al. 2011; Melnyk et al. 2009; Noble et al. 2006). Previous research has found that when shopping in physical stores, women are more likely to browse products and use the time for social interaction with others (e.g., merchants). However, men are more needs-driven. They are less likely to spend time browsing various products or interacting with others (Melnyk et al. 2009; Noble et al. 2006). In addition, social interaction has been argued to positively influence consumer loyalty (Melnyk et al. 2009; Noble et al. 2006). Women have been found to have higher loyalty to individual store employees; while men are more likely to be loyal to a firm as whole (Melnyk et al. 2009). Future research can investigate whether such differences hold in the virtual world. If so, virtual malls may need to develop strategies and customer service policies by considering such gender differences. For example, it could be better to have personalized sales representative avatars for female customers and try to build a personal relationship with them, while this may be not be necessary for male customers. To attract female customers, it could be a good idea to open localized virtual stores with personalized services. In order to attract male customers, relatively large virtual chain stores might be better. Previous research has also found that older people and women tended to spend more time browsing e-commerce websites, (Danaher et al. 2006). In the virtual world, different strategies for stores targeting different age groups are desired and possible.

The study site of this research is Second Life, a creativity-oriented electronic environment that mimics the physical world, where people can interact with each other through their avatars and create various types of virtual objects (Bainbridge 2007). In addition to Second Life, there is another popular type of virtual world, massively multiplayer online games (MMOGs), where players typically try to increase the abilities of their avatars and work with other players to complete a goal (Williams et al. 2009). Based on the characteristics of these two types of virtual worlds, previous research has suggested that when studying social psychology or cognitive science related phenomena, it would be better to use Second Life, since it mimics the physical world and thus the results could be generalized to the physical world; however, MMOGs can be leveraged to study social interactions and economic systems, since players need to work with team members to achieve a common goal and purchase items to make their avatars stronger (Bainbridge 2007). This study examined avatar social interactions based on gender and age differences in Second Life. Future research can further investigate social interaction patterns among team members in MMOGs. In addition to individual level social interactions, future research can also examine the group-level interactions among different teams in MMOGs and their success in completing their goals.

This study has several other limitations which future research needs to address. First, avatar interactions were modeled as undirected graphs. Future research can explore how to model them as a directed graph by capturing who is seeking help from whom in an interaction. Future research may also consider adding interaction intensity for each edge in the network. The duration of each interaction could be used to indicate intensity. Second, in this study, I used dichotomous categories of virtual age. Further research can consider using a finer granularity.

CHAPTER 6. CONCLUSIONS AND FUTURE DIRECTIONS

6.1 Contributions

The rapid development and evolution of the Internet have enabled people to access information whenever and wherever they want. Recently, with the advent of Web 2.0, the Internet has evolved towards multimedia-rich content delivery, end-user content generation, and community-based social interaction (O'Reilly 2005). As a result, the amount of social media content has grown tremendously, providing valuable resources for analyzing and understanding various social phenomena in the new media. Different from the world where people physically live, the new media bring additional types of worlds into people's lives: online worlds and virtual worlds. Examples of online worlds include Web forums, blogs, and online reviews, while the most famous example of virtual worlds is Second Life. My dissertation aims to address the overarching question about how people adapt to social media to share information and exchange opinions, and what factors influence their activities in the new media. I adopt Web mining, machine learning, and computational linguistics techniques to analyze aspects of people and their behavior, such as gender differences, emotional differences, avatar activity differences, and avatar social interaction differences in online and virtual worlds. The guiding theories for the analyses include systematic functional linguistic theory, social information processing theory, social presence theory, (social) role theory, and gender role theory.

Chapter 2 uses feature-based online social media text classification techniques to investigate online gender differences between female and male participants in Web forums, by examining their writing styles and topics of interest. The proposed feature-based gender classification framework is generic and can be applied to different Web forums. In the framework, different types of features are examined, including: lexical features, syntactic features, structural features, and content-specific features. This study has made several contributions. First, a systematic framework of gender classification is proposed to analyze online gender differences in social media, an area which has received little investigative attention. The framework can be applied to study gender differences in many other domains (e.g., sociology, business, and marketing). It provides an informative point of departure for continued research. Second, the empirical study demonstrated the effectiveness of the proposed framework, thus confirming the prevalence of online gender differences in Web forums. Third, this study also makes a research contribution by examining different feature sets and identifying the one with the best classification performance. The comparison of different feature sets also indicates the importance of incorporating content-specific features and conducting feature selection in automatic gender classification for online social media.

Chapter 3 investigates the emotional differences between the two genders in text-based communications in the online world. An automatic emotion detection framework using sentiment analysis techniques is developed. This study has made several contributions. First, unlike most previous research, the proposed framework takes into account longitudinal textual data from an entire social media site to conduct the emotional difference analysis automatically, which can provide more comprehensive and unbiased analysis results. Specifically, different algorithms are developed to analyze the sentence-level subjectivity and phrase- and word-level polarity, and then conduct statistical analysis to compare the differences between men and women. To the best of my knowledge, this is the first study to utilize advanced sentiment analysis techniques to analyze emotional differences between the two genders. In addition, no previous studies have specifically examined these differences in Web forum communication. Using the proposed framework, an empirical experiment is conducted on a large and long-standing international women's political forum with more than 30,000 messages. The sentence-level analysis results indicated that women were significantly more subjective than men. The phrase- and word-level analysis results showed that in general women were significantly more likely to express both positive and negative emotions as compared to men. By investigating the emotional content generated by women, the existence of discussions related to stereotyping and social roles was observed.

Chapter 4 examines whether real-life social norms hold in the virtual world and what major factors can influence people's behavior in the virtual world. Guided by the theories of social presence, social role, and gender role, the effects of avatar virtual gender, virtual age, and region theme on avatars' physical activities are examined. This study has significant research contributions to both the technical and behavioral research fields of IS. For technical research, this study developed an integrated framework for data collection in the virtual world. Currently, for researchers who are interested in virtual worlds, data collection is a critical problem. It is always difficult to collect avatar related data since most virtual worlds do not provide easy data collection function. The integrated data collection framework developed in this study provides a systematic way for collecting various types of avatar behavior and profile data. In addition, the improved bot-based approach in the framework has addressed the limitations of the existing bot-based approach developed by Yee et al. (2007). By utilizing the LibOpenMetaverse, the improved bot-based approach allows each bot to conduct a much larger-scale data collection in a much faster manner. Further, by combining the improved bot-based approach with a spider-based approach, the framework is able to collect not only avatar behavioral data but also avatar profile data, which hadn’t been done in previous research.

For behavioral research, this study examines and identifies similar gender and age differences toward activities in the virtual world compared to the physical world. Similarly to the physical world, male avatars are more (less) likely to perform high-active (low-active) actions than female avatars, while young-aged avatars are more (less) likely to perform high-active (low-active) actions than old-aged avatars. In addition, this study introduces region theme as an important contextual factor for the virtual world, and shows that region theme also has a significant impact on avatar activity. Specially, avatars in commercial transaction regions are more (less) likely to perform high-active (low-active) actions than avatars in non-profit, help-supporting regions. Furthermore, the study examines both the main and interaction effects of virtual gender, virtual age, and region theme on avatars’ physical activities, which has not been examined before.

Chapter 5 further explores avatars’ gender and age differences in their help-seeking interactions in the virtual world. Avatar social interaction related data is collected and calculated. This study has made several contributions to the virtual world research. First, although previous research has studied gender and age differences on people's help-seeking behavior in real-world settings, no study has examined such differences in virtual worlds. Previous virtual world studies examined gender differences in avatar interpersonal distances and eye contact patterns, and the impact of both gender and age in avatar physical activity differences. But no study has examined the gender and age differences in avatar social interactions. This study has further tested gender role theory by examining avatar gender and age differences in their social interactions in help-seeking regions in the virtual world. Second, systematic analyses are conducted to examine such differences. Both the overall gender and age difference analyses and detailed investigations are conducted. In the detailed analyses, three types of interaction networks based on gender or age are compared. The three types of networks based on gender are male-male, female-female, and male-female/female-male interaction networks. The three types of networks based on age are young-young, old-old, and young-old/old-young interaction networks.

The investigation results showed that old-aged avatars have larger degree and higher betweenness centrality, HITS authority/hub, and RageRank scores than young-aged avatars, indicating that avatar are more likely to interact with old-aged avatars to seek help, and old-aged avatars tend to be in the more important positions in the interaction network. Results of the detailed avatar gender difference analysis indicate that female avatars have more interactions with female avatars when seeking help compared with interactions among male avatars. In addition, both male-male and female-female interaction networks had significantly higher betweenness centrality, HITS, and PageRank scores than the male-female/female-male interaction network, indicating that the interaction networks of the same gender tended to be more centralized than the network of interactions between the two genders. Results of the detailed avatar age difference analysis show that the old-old interaction network had the largest degree, followed by the young-old/old-young interaction network and the young-young interaction network, suggesting that when seeking help, both old-aged and young-aged avatars were more likely to interact with old-aged avatars. In addition, both young-young and old-old interaction networks had significantly higher betweenness centrality, HITS, and PageRank scores than the young-old/old-young interaction network, indicating that on average the interaction networks of the same age group tend to be more centralized than the interaction network of the two different age groups.

Overall, the research studies in my dissertation contribute to the literature in social media analytics, knowledge discovery, virtual world research, and text and Web mining.

6.2 Relevance to Management Information Systems Research

Two important and complementary research paradigms in Information Systems (IS) are design science and behavioral science paradigms (Hevner et al. 2004). The design science paradigm focuses on creative exploration and experimentation of novel ideas, modeling methods, analytical techniques, computational algorithms, or visualization designs that can be instantiated (e.g., implemented in a technical research framework) for significant gains in effectiveness or efficiency (Denning 1997; Hevner et al. 2004). According to the three-cycle view by Hevner (Hevner 2007) the design science paradigm embraces relevance, rigor, and design, that together allow researchers and practitioners to generate IT artifacts capable of generating substantially improved utility for targeted problems. Different from the design science paradigm, the behavioral science paradigm emphasizes the development, justification, or testing of theory related to organizational and human phenomena surrounding the analysis, design, implementation, management, and use of information systems (Hevner et al. 2004).

The studies in my dissertation are relevant to both paradigms in IS research. Specifically, different IT artifacts are developed as technical frameworks and algorithms. The creation of these IT artifacts belongs to the design science paradigm. In addition, various social psychology theories are leveraged to guide the hypothesis development and data analyses. This follows the line of the behavioral science paradigm.

Focusing on the online world, the feature-based gender classification framework developed in Chapter 2 and the integrated emotion detection framework with sentence-level, phrase-level, and word-level emotion detection algorithms created in Chapter 3 are the targeted IT artifacts. Experimental analyses have demonstrated the effectiveness of these IT artifacts. Stereotyping and social roles theories are used to guide the data analyses. The results have provided evidence of the existence of gender and emotional differences in online communications. Similar gender and emotional patterns have been observed towards the online world compared to the physical world.

Focusing on the virtual world, the combined bot-based and spider-based avatar data collection framework presented and adopted in Chapters 4 and 5 is the IT artifact. It enables conducting a much larger-scale data collection in a much faster manner. This new technical framework has significant contribution to the design science paradigm. Guided by the theories of social presence, social role, and gender role, detailed analyses are conducted to examine the effects of avatars' virtual gender and virtual age (as well as region theme for physical activities) on avatars' physical activities and their help-seeking interactions in the virtual world. As exploratory studies, some interesting patterns have been observed, which provide better understanding of the validity of the leveraged social psychology theories in the virtual world settings.

Overall, my dissertation studies are relevant to both the design science and behavioral science paradigms in IS research by developing technical frameworks and algorithms as IT artifacts and leveraging social psychology theories to guide the hypothesis development and data analyses.

6.3 Future Research Directions

The research studies in my dissertation have examined aspects of people and their behavior, such as gender differences, emotional differences, avatar activity differences, and avatar social interaction differences in online and virtual worlds. Chapters 2 and 3 focus on analyzing the gender differences and emotional differences in online worlds. Research frameworks are developed based on gender classification techniques and emotion detection algorithms. In both studies, the empirical experiments were conducted on Web forum content. Future research can adopt the proposed frameworks to examine the gender and emotional differences in other communication media in the online world, such as blogs, social networking sites, and wikis. Future research can also extend the proposed research frameworks to other languages. Additional language translation component and a scalable multilingual feature representation and extraction method need to be developed. In additoin, future research can apply the proposed frameworks to further investigate various important social media domains, such as marketing, e-commerce, health care, and education, to obtained detialed understanding of doman-specific gender and emotional differences. Furthermore, the types of emotions measured in Chapter 3 focus on positivity versus negativity. Future research can conduct fine-granularity analysis to examine more specific types of emotions, such as happiness, love, life satisfaction, etc., versus fear, sadness, anger, etc.

Chapters 4 and 5 focus on investigating avatar activity differences and avatar social interaction differences in virtual worlds. Both studies use the nonverbal data to conduct analyses. Future research can examine how to collect verbal data from the virtual world and how to leverage both verbal and nonverbal data together to study avatar behavior and interactions. In addition to avatars' physical activities, future research can also investigate the impacts of avatar gender, age, and region theme on other types of avatar behavior, such as sentiments, emotions, and personality (positivity, friendliness, altruism, etc.). Furthermore, the gender and age information used in both studies is the virtual gender and virtual age of avatars instead of the real gender and real age of people who control those avatars. Future research needs to link an avatar's virtual gender and age to the real gender and age of the person whom the avatar represents to further investigate and compare gender and age impacts on behavior in the virtual and physical worlds. In terms of avatar social interactions, Chapter 5 models the interactions as undirected graphs. Future research can explore how to model them as a directed graph by capturing who is seeking help from whom in an interaction. Future research may also consider adding interaction intensity for each edge in the network. In addition to the help-seeking scenario, future research needs explore the unique characteristics of avatar social interactions in other areas such as marketing. For example, future research can examine whether gender-based social interaction differences in people’s shopping behavior in the physical world exist in the virtual worlds, which could help develop marketing strategies in the virtual world. Furthermore, the study site of both chapters is Second Life, a popular virtual world that mimics the physical world. Future research can explore avatar gender, age, and social interaction differences in MMOGs, another type of popular virtual world, where players typically try to increase the abilities of their avatars and work with other players to complete a goal. A great deal of social interactions occur in MMOGs, since players need to work with team members to achieve a common goal and purchase items to make their avatars stronger. Future research can examine different social interaction patterns, such as within-team and between-team interactions, in MMOGs.

REFERENCES

Abbasi, A., and Chen, H. 2005. "Applying Authorship Analysis to Extremist-Group Web Forum Messages," *IEEE Intelligent Systems* (20:5), pp 67-75.

Abbasi, A., and Chen, H. 2008. "Writeprints: A Stylometric Approach to Identity- Level Identification and Similarity Dectection in Cyberspace," *ACM Transactions on Information Systems* (26:2), pp 1-29.

Abbasi, A., Chen, H., and Nunamaker, J.F. 2008a. "Stylometric Identification in Electronic Markets: Scalability and Robustness," *Journal of Management Information Systems* (25:1), pp 49-78.

Abbasi, A., Chen, H., and Salem, A. 2008b. "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," *ACM Transactions on Information Systems* (26:3), pp 1-34.

Antonucci, T.C., and Akiyama, H. 1987. "An Examination of Sex Differences in Social Support among Older Men and Women," *Sex Roles* (77:11/12), pp 737-749.

Argamon, S., Koppel, M., and Avneri, G. 1988. "Routing Documents According to Style.," *Proceedings of the 1st International Workshop on Innovative Information*, Pisa, Italy.

Argamon, S., Koppel, M., Fine, J., and Shimoni, A. 2003a. "Gender, Genre, and Writing Style in Formal Written Texts," *Text* (23:3), pp 321-346.

Argamon, S., Saric, M., and Stein, S.S. 2003b. "Style Mining of Electronic Messages for Multiple Authorship Discrimination," *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 475-480.

Awad, N.F., and Ragowsky, A. 2008. "Establishing Trust in Electronic Commerce through Online Word of Mouth: An Examination across Genders," *Journal of Management Information Systems* (24:4), pp 101-121

Baayen, R.H., Halteren, H.V., Neijt, A., and Tweedie, F.J. 2002. "An Experiment in Authorship Attribution," *Proceedings of the 6th International Conference on Statistical Analysis of Textual Data*, pp. 69-75.

Baayen, R.H., Halteren, H.V., and Tweedie, F.J. 1996. "Outside the Cave of Shadows: Using Syntactic Annotation to Enhance Authorship Attribution," *Literary and Linguistic Computing* (11:3), pp 121-132.

Bainbridge, W.S. 2007. "The Scientific Research Potential of Virtual Worlds," *Science* (317), pp 472-476.

BBC. 2009. "Females Less Physically Active," in: *BBC News*.

Belcher, B.R., Berrigan, D., Dodd, K.W., Emken, B.A., Chou, C.-P., and Spuijt-Metz, D. 2010. "Physical Activity in Us Youth: Impact of Race/Ethnicity, Age, Gender, and Weight Status," *Medicine and Science in Sports and Exercise* (doi: 10.1249/MSS.0b013e3181e1fba9).

Benamara, F., Cesarano, C., Picariello, A., Reforgiato, D., and Subrahmanian, V.S. 2007. "Sentiment Analysis: Adjectives and Adverbs Are Better Than Adjectives Alone," *the International Conference on Weblogs and Social Media (ICWSM-2007)*, Boulder, CO: AAAI Press, pp. 203-206.

Biddle, B.J. 1986. "Recent Development in Role Theory," *Annual Review of Sociology*), pp 1267-1292.

Bimber, B. 2000. "Measuring the Gender Gap on the Internet," *Social Science Quarterly* (81:3), pp 868-876.

Blascovich, J., Loomis, J., Beall, A.C., Swinth, K.R., Hoyt, C.L., and Bailenson, J.N. 2002. "Immersive Virtual Environment Technology as a Methodological Tool for Social Psychology," *Psychological Inquiry* (13:2), pp 103-124.

Bowling, A. 1991. "Social Support and Social Networks: Their Relationship to the Successful and Unsuccessful Survival of Elderly People in the Community. An Analysis of Concepts and a Review of the Evidence," *Journal of Family Practice* (8:1), pp 68-83.

Brebner, J. 2003. "Gender and Emotions," *Personality and Individual Differences* (34), pp 387-394.

Brin, S., and Page, L. 1998. "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems* (30:1-7), pp 107-117.

Briton, N.J., and Hall, J.A. 1995. "Gender-Based Expectancies and Observer Judgments of Smiling," *Journal of Nonverbal Behavior* (19), pp 49-65.

Brody, L., and Hall, J. 2008. "Gender and Emotion in Context," in: *Handbook of Emotions (3 Ed)*, M. Lewis, J. Haviland-Jones and L.F. Barrett (eds.). New York: The Guildford Press, pp. 395-408.

Brody, L.R. 1985. "Gender Differences in Emotional Development: A Review of Theories and Research," *Journal of Personality* (53:2), pp 102-149.

Brody, L.R. 1997. "Gender and Emotion: Beyond Stereotypes," *Journal of Social Issues* (53:2), pp 369-394.

Brody, L.R., and Hall, J. 1993. "Gender and Emotion," in: *Handbook of Emotions*, M. Lewis and J. Haviland (eds.). New York: Guilford Press, pp. 447-461.

Cameron, D. 2007. *The Myth of Mars and Venus: Do Men and Women Really Speak Different Languages?* Oxford: Oxford University Press.

Caspersen, C.J., Pereira, M.A., and Curran, K.M. 2000. "Changes in Physical Activity Patterns in the United States, by Sex and Cross-Sectional Age," *Medicine and Science in Sports and Exercise* (32:9), pp 1601-1609.

Chaovalit, P., and Zhou, L. 2005. "Movie Review Mining: A Comparison between Supervised and Unsupervised Classification Approaches," *the 38th Annual Hawaii International Conference on System Sciences*, Hawaii, HI: IEEE Press.

Chen, H. 2009. "Ai, E-Government, and Politics 2.0," *IEEE Intelligent Systems* (24:5), pp 64-67.

Chodorow, N.J. 1994. *Feminities, Masculinities, Sexualities: Freud and Beyond*. Lexington, Kentucky: University Press of Kentucky.

Coates, J. 1993. *Women, Men and Language (2nd Ed.)*. New York: Longman Inc.

CommerceNet. 1999. "The Commercenet/Nielsen Internet Demographic Survey (1999)," http://www.commerce.net/).

Consaluo, M., and Paasonen, S. 2002. *Women & Everyday Uses of the Internet: Agency & Identity*. New York: Peter Lang Publishing.

Corney, M., deVel, O., Anderson, A., and Mohay, G. 2002. "Gender-Preferential Text Mining of E-Mail Discourse," *Proceedings of the 18th Annual Computer Security Applications Conference (ACSAC 2002)*, Las Vegas, pp. 282–292.

Danaher, P.J., Mullarkey, G.W., and Essegaier, S. 2006. "Factors Affecting Web Site Visit Duration: A Cross-Domain Analysis," *Journal of Marketing Research* (43:2), pp 182-194.

Dang, Y., Zhang, Y., and Chen, H. 2010. "A Lexicon Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews," *IEEE Intelligent Systems* (25:4), pp 46-53.

Dash, M., and Liu, H. 1997. "Feature Selection for Classification," *Intell. Data Anal.* (1:3), pp 131-156.

Dave, K., Lawrence, S., and Pennock, D. 2003. "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews," *the International World Wide Web Conference*, Budapest, Hungary: ACM Press, pp. 519–528.

Denecke, K. 2008. "Using Sentiwordnet for Multilingual Sentiment Analysis," *the IEEE 24th International Conference on Data Engineering Workshop (ICDEW 2008)*, Cancun: IEEE Press, pp. 507-512.

Denning, P.J. 1997. "A New Social Contract for Research," *Communications of the ACM* (40:2), pp 132-134.

deVel, O. 2000. "Mining E-Mail Authorship," *Paper presented at the Workshop on Text Mining, ACM International Conference on Knowledge Discovery and Data Mining (KDD 2000)*, Boston, MA.

deVel, O., Anderson, A., Corney, M., and Mohay, G. 2001. "Mining E-Mail Content for Author Identification Forensics," *SIGMOD Record* (30:4), pp 55–64.

Devitt, A., and Ahmad, K. 2007. "Sentiment Polarity Identification in Financial News: A Cohesion-Based Approach," *the 45th Annual Meeting of the Association of Computational Linguistics*, Prague: ACL Press, pp. 984-991.

Diederich, J., Kindermann, J., Leopold, E., and Paass, G. 2003. "Authorship Attribution with Support Vector Machines," *Applied Intelligence* (19:1-2), pp 109–123.

Diversified Media Design, Combinedstory, and Market Truths Limited. 2007. "The Virtual Brand Footprint: The Marketing Opportunity in Second Life,").

Eagly, A., and Karau, S. 1991. "Gender and the Emergence of Leaders: A Meta-Analysis," *Journal of Personality and Social Psychology* (60:5), pp 685-710.

Eagly, A.H. 1987. *Sex Differences in Social Behavior: A Social Role Interpretation*. Hillsdale, NJ: Erlbaum.

Eagly, A.H., and Wood, W. 1991. "Explaining Sex Differences in Social Behavior: A Meta-Analytic Perspective," *Personality and Social Psychology Bulletin* (17), pp 306-315.

Esuli, A., and Sebastiani, F. 2006. "Sentiwordnet: A Publicly Available Lexical Resource for Opinion Mining," *LREC-06, the 5th Conference on Language Resources and Evaluation*, Genova, Italy, pp., 417-422.

Fabes, R.A., and Martin, C.L. 1991. "Gender and Age Stereotypes of Emotionality," *Personality and Social Psychology Bulletin* (17), pp 532-540.

Fahrni, A., and Klenner, M. 2008. "Old Wine or Warm Beer: Target-Specific Sentiment Analysis of Adjectives," *Proceedings of the AISB 2008 Symposium on Affective*

Language in Human and Machine, Aberdeen, Scotland: The Society for the Study of Artificial Intelligence and Simulation of Behaviour Press, pp. 60-63.

Fei, Z., Liu, J., and Wu, G. 2004. "Sentiment Classification Using Phrase Patterns," *the 4th IEEE International Conference on Computer Information Technology*: IEEE Press, pp. 1147-1152.

Fischer, A.H., and Manstead, A.S.R. 2004. "Gender and Culture Differences in Emotion," *Emotion* (4:1), pp 87-94.

Fischer, E., and Arnold, S.J. 1990. "More Than a Labor of Love: Gender Roles and Christmas Gift Shopping," *Journal of Consumer Research* (17:3), pp 333-345.

Fiske, S., and Stevens, L. 1993. "What's So Special About Sex? Gender Stereotyping and Discrimination," in: *Gender Issues in Contemporary Society*, S. Oskamp and M. Costanzo (eds.). Newbury Park, CA: Sage Publications, pp. 173-196.

Forman, G. 2003. "An Extensive Empirical Study of Feature Selection Metrics for Text Classification," *The Journal of Machine Learning Research* (3), pp 1289-1305.

Forsyth, R.S., and Holmes, D.I. 1996. "Feature Finding for Text Classification," *Literary and Linguistic Computing* (11:4), pp 163–174.

Fountain, J.E. 2000. "Constructing the Information Society: Women, Information Technology, and Design," *Technology and Society* (22:1), pp 45-62.

Fox, E. 2008. *Emotion Science*. Basingstoke: Palgrave Macmillan.

Franceschi, K., Lee, R.M., Zanakis, S.H., and Hinds, D. 2009. "Engaging Group E-Learning in Virtual Worlds," *Journal of Management Information Systems* (26:1), pp 73-100.

Freeman, L. 1977. "A Set of Measures of Centrality Based on Betweenness," *Sociometry* (40), pp 35-41.

Freeman, L. 1979. "Centrality in Social Networks: Conceptual Clarification," *Social Networks* (1), pp 215-239.

Fujita, F., Diener, E., and Sandvik, E. 1991. "Gender Differences in Negative Affect and Well-Being: The Case for Emotional Intensity," *Journal of Personality and Social Psychology* (61), pp 427-434.

Fuller, J.E. 2004. "Equality in Cyberdemocracy? Gauging Gender Gaps in on-Line Civic Participation," *Social Science Quarterly* (85:4), pp 938-957.

Gamon, M. 2004. "Sentiment Classification on Customer Feedback Data: Noisy Data, Large Feature Vectors, and the Role of Linguistic Analysis," *Proceedings of the 20th International Conference on Computational Linguistics*, Geneva, CH: ACL Press, pp. 841-847.

Gamon, M., Aue, A., Corston-Oliver, S., and Ringger, E. 2005. "Pulse: Mining Customer Opinions from Free Text," *the 6th International Symposium on Intelligent Data Analysis*, Madrid, Spain.

Gordon-Larsen, P., McMurray, R.G., and Popkin, B.M. 2000. "Determinants of Adolescent Physical Activity and Inactivity Patterns," *Pediatrics* (105:6), p e83.

Gordon, R., Björklund, N.K., Smith, R.J., and Blyden, E.R. 2009. "Halting Hiv/Aids with Avatars and Havatars: A Virtual World Approach to Modelling Epidemics," *BMC Public Health* (9:S13), pp 1-6.

Gove, W.R. 1978. "Sex Differences in Mental Illness among Adult Men and Women: An Evaluation of Four Questions Raised Regarding the Evidence on the Higher Rates of Women," *Social Science and Medicine* (12), pp 187-198.

Gray, J. 1992. *Men Are from Mars, Women Are from Venus*. New York: HarperCollins.

Grefenstette, G., Qu, Y., Shanahan, J.G., and Evans, D.A. 2004. "Coupling Niche Browsers and Affect Analysis for an Opinion Mining Application," *Proceedings of the 12th International Conference Recherche d'Information Assistee par Ordinateur*, pp. 186-194.

Grinstein-Weiss, M., Fishman, G., and Eisikovits, Z. 2005. "Gender and Ethnic Differences in Formal and Informal Help Seeking among Israeli Adolescents," *Journal of Adolescence* (28:6), pp 765-779.

Grossman, M., and Wood, W. 1993. "Sex Differences in the Intensity of Emotional Experience: A Social Role Interpretation," *Journal of Personality and Social Psychology* (65), pp 1010-1022.

Guiller, J., and Durndell, A. 2007. "Students' Linguistic Behaviour in Online Discussion Groups: Does Gender Matter? ," *Computers in Human Behavior* (23:5), pp 2240-2255.

Guo, B., and Nixon, M.S. 2009. "Gait Feature Subset Selection by Mutual Information," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* (39:1), pp 36-46.

Halbert, D. 2004. "Shulamith Firestone: Radical Feminism and Visions of the Information Society," *Information Communication and Society* (7:1), pp 115-136.

Halliday, M.A.K. 1985. *An Introduction to Functional Grammar*, (1 ed.). London: Arnold.

Halliday, M.A.K., and Matthiessen, C.M.I.M. 2004. *An Introduction to Functional Grammar*, (3 ed.). London: Arnold.

Harasim, L. 1990. "Introduction to Online Education," in: *Online Education: Perspectives on a New Environment*, L. Harasim (ed.). New York: Praeger Publishers, pp. 17-23.

Harcourt, W. 2000. "The Personal and the Political: Women Using the Internet," *CyberPsychology and Behavior* (3:5), pp 693-697.

Harp, D., and Tremayne, M. 2006. "The Gendered Blogosphere: Examining Inequality Using Network and Feminist Theory," *Journalism and Mass Communication Quarterly* (83:2), pp 247-264.

Harris, H., Bailenson, J.N., Nielsen, A., and Yee, N. 2009. "The Evolution of Social Behavior over Time in Second Life," *Presence: Teleoperators and Virtual Environments* (18:6), pp 434-448.

Hassell, M., Goyal, S., Limayem, M., and Boughzala, I. 2009. "Being There: An Empirical Look at Learning Outcomes in 3d Virtual Worlds," *15th Americas Conference on Information Systems (AMCIS)*, San Francisco, California, pp. 1-11, Paper 733.

Hatzivassiloglou, V., and Wiebe., J. 2000. "Effects of Adjective Orientation and Gradability on Sentence Subjectivity," *the 18th International Conference on Computational Linguistics*: ACL Press, pp. 299-305.

Hemp, P. 2006. "Avatar-Based Marketing," *Harvard Business Review* (84:6), pp 48-56.

Hendaoui, A., Limayem, M., and Thompson, C.W. 2008. "3d Social Virtual Worlds: Research Issues and Challenges," *IEEE Internet Computing* (12:1), pp 88-92.

Hevner, A.R. 2007. "A Three Cycle View of Design Science Research," *Scandinavian Journal of Information Systems* (19:2), pp 87-92.

Hevner, A.R., March, S.T., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp 75-105.

Hindin, M.J. 2007. "Role Theory," in: *The Blackwell Encyclopedia of Sociology*, G. Ritzer (ed.). Blackwell Publishing, pp. 3959-3962.

Holzwarth, M., Janiszewski, C., and Neumann, M.M. 2006. "The Influence of Avatars on Online Consumer Shopping Behavior," *Journal of Marketing* (70:4), pp 19-36.

Hota, S., Argamon, S., Koppel, M., and Zigdon, I. 2006. "Performing Gender: Automatic Stylistic Analysis of Shakespeare's Characters," *Proceedings of the Digital Humanities Conference (Association for Computers in Humanities and the Association for Literary and Linguistic Computing)*, pp. 100-106.

Hovy, S.-M.K.a.E. 2006. "Automatic Identification of Pro and Con Reasons in Online Reviews," *the COLING/ACL Main Conference* Morristown, NJ: ACL Press, pp. 483-490.

Hu, M., and Liu, B. 2004. "Mining and Summarizing Customer Reviews," *the ACM SIGKDD International Conference*: ACM Press, pp. 168-177.

Hu, P.J.-H., Cheng, T.-H., Wei, C.-P., Yu, C.-H., Chan, A.L.F., and Wang, H.-Y. 2007. "Managing Clinical Use of High-Alert Drugs: A Supervised Learning Approach to Pharmacokinetic Data Analysis," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* (37:4), pp 481-492.

Jackson, L.A., Ervin, K.S., Gardner, P.D., and Schmitt, N. 2001. "Gender and the Internet: Women Communicating and Men Searching," *Sex Roles: A Journal of Research* (44:5-6), pp 363-378.

Jackson, V., Stoel, L., and Brantley, A. 2011. "Mall Attributes and Shopping Value: Differences by Gender and Generational Cohort," *Journal of Retailing and Consumer Services* (18), pp 1-9.

Jaffe, J.M., Lee, Y.E., Huang, L., and Oshagan, H. 1999. "Gender Identification, Interdependence and Pseudonyms in Cmc: Language Patterns in an Electronic Conference," *The Information Society* (15), pp 221-234.

Kawale, J., Pal, A., and Srivastava, J. 2009. "Churn Prediction in Mmorpgs: A Social Influence Based Approach," *Proceedings of the IEEE International conference on Computational Sciences and Engineering*: IEEE Computer Society, pp. 423-428.

Keegan, B., Ahmad, M., Srivastava, J., Williams, D., and Contractor, N. 2010. "Dark Gold: Statistical Properties of Clandestine Networks in Massively Multiplayer Online Games," *Proceedings of IEEE 2nd international conference Social Computing*: IEEE Computer Society, pp. 201-208.

Kehrwald, B. 2008. "Understanding Social Presence in Text-Based Online Learning Environments," *Distance Education* (29:1), pp 89-106.

Kidder, D. 2002. "The Influence of Gender on the Performance of Organizational Citizenship Behaviors," *Journal of Management Information Systems* (28:5), pp 629-648.

Kleinberg, J.M. 1999. "Authoritative Sources in a Hyperlinked Environment," *Journal of the ACM* (46:5), pp 604-632.

Koppel, M., Akiva, N., and Dagan, I. 2006. "Feature Instability as a Criterion for Selecting Potential Style Markers," *J. Amer. Soc. Inf. Sci. Technol* (57:11), pp 1519–1525.

Koppel, M., Argamon, S., and Shimoni, A. 2002. "Automatically Categorizing Written Texts by Author Gender," *Literary and Linguistic Computing* (14:7), pp 401-412.

Koppel, M., and Schler, J. 2003. "Exploiting Stylistic Idiosyncrasies for Authorship Attribution," *Proceedings of the IJCAIWorkshop on Computational Approaches to Style Analysis and Synthesis*, Acapulco, Mexico.

Landman-Peeters, K.M.C., Hartman, C.A., Pompe, G.v.d., Boer, J.A.d., Minderaa, R.B., and Ormel, J. 2005. "Gender Differences in the Relation between Social Support, Problems in Parent-Offspring Communication, and Depression and Anxiety," *Social Science & Medicine* (60:11), pp 2549-2559.

Ledger, G.R., and Merriam, T.V.N. 1994. "Shakespeare, Fletcher, and the Two Noble Kinsmen.," *Literary and Linguistic Computing* (9:4), pp 235–248.

Li, J., Su, H., Chen, H., and Futscher, B.W. 2007. "Optimal Search-Based Gene Subset Selection for Gene Array Cancer Classification," *IEEE Transactions on Information Technology in Biomedicine* (11:4), pp 398-405.

Li, J., Zhang, Z., Li, X., and Chen, H. 2008. "Kernel-Based Learning for Biomedical Relation Extraction," *Journal of the American Society for Information Science and Technology (JASIST)* (59:5), pp 756-769.

Li, J., Zheng, R., and Chen, H. 2006. "From Fingerprint to Writeprint," *Communications of the ACM* (49:4), pp 76-82.

Liu, B. 2007. *Web Data Mining*. Berlin Heidelberg: Springer.

Lucas, R.E., and Gohm, C.L. 2000. "Age and Sex Differences in Subjective Well-Being across Cultures," in: *Culture and Subjective Wellbeing*, E. Diener and E.M. Suh (eds.). Cambridge, MA: MIT Press, pp. 291-318.

Mahmoodi, H., and Jalali, A.A. 2009. "Virtual Age: Enabling Technologies and Trends," *Sixth International Conference on Information Technology: New Generations*, pp. 999-1004.

Malcom, N. 2003. "Constructing Female Athleticism: A Study of Girls' Recreational Softball," *American Behavioral Scientist* (46:10), pp 1387-1404.

Martindale, C., and Mckenzie, D. 1995. "On the Utility of Content Analysis in Author Attribution: The Federalist," *Comput. Humanit.* (29:4), pp 259-270.

Mauss, I.B., and Robinson, M.D. 2009. "Measures of Emotion: A Review," *Cognition and Emotion* (23:2), pp 209-237.

Meiri, R., and Zahavi, J. 2006. "Using Simulated Annealing to Optimize the Feature Selection Problem in Marketing Applications," *European Journal of Operational Research* (171:3), pp 842-858.

Melnyk, V., Osselaer, S.M.J.v., and Bijmolt, T.H.A. 2009. "Are Women More Loyal Customers Than Men? Gender Differences in Loyalty to Firms and Individual Service Providers," *Journal of Marketing* (73:4), pp 82-96.

Mendenhall, T.C. 1887. "The Characteristic Curves of Composition," *Science* (11:11), pp 237–249.

Messinger, P.R., Stroulia, E., Lyons, K., Bone, M., Niu, R.H., Smirnov, K., and Perelgut, S. 2009. "Virtual Worlds - Past, Present, and Future: New Directions in Social Computing," *Decision Support Systems* (47:3), pp 204-228.

Mihalcea, R. 2004. "Graph-Based Ranking Algorithms for Sentence Extraction, Applied to Text Summarization," *In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL 2004)*, Barcelona, Spain: ACL.

Miller, J.B. 1976. *Toward a New Psychology of Women*. Boston: Beacon Press.

Mishne, G. 2005. "Experiments with Mood Classification in Blog Posts," *the 1st Workshop on Stylistic Analysis of Text for Information Access, at SIGIR 2005*, Salvador, Brazil: ACM Press.

Mitra, A. 2004. "Voices of the Marginalized on the Internet: Examples from a Website for Women of South Asia," *Journal of Communication* (54:3), pp 492-510.

National Election Study. 1998. "American National Election Study. 1998 Pre- and Post-Election Survey," *Conducted by the Center for Political Studies of the Institute for Social Research, The University of Michigan, Ann Arbor, Inter-University Consortium for Political and Social Research*).

Newman, M.L., Groom, C.J., Handelman, L.D., and Pennebaker, J.W. 2008. "Gender Differences in Language Use: An Analysis of 14,000 Text Samples," *Discourse Processes* (45), pp 211-236.

Noble, S.M., Griffith, D.A., and Adjei, M.T. 2006. "Drivers of Local Merchant Loyalty: Understanding the Influence of Gender and Shopping Motives," *Journal of Retailing* (82:3), pp 177-188.

Nolen-Hoeksema, S. 1987. "Sex Differences in Unipolar Depression: Evidence and Theory," *Psychological Bulletin* (101), pp 259-282.

Nowson, S., and Oberlander, J. 2006. "The Identity of Bloggers: Openness and Gender in Personal Weblogs," *Proceedings of the AAAI Spring Symposia on Computational Approaches to Analyzing Weblogs*, Stanford, California.

O'Madadhain, J., Fisher, D., Smyth, P., White, S., and Boey, Y. 2005. "Analysis and Visualization of Network Data Using Jung," *Journal of Statistical Software* (10:2), pp 1-25.

O'Reilly, T. 2005. "What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software," http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html).

Ogan, C., Cicek, F., and Ozakca, M. 2005. "Letters to Sarah: Analysis of Email Responses to an Online Editorial," *New Media and Society* (7:4), pp 533-557.

Palapattu, A., Kinsgery, J., and Ginsburg, G. 2006. "Gender Role Orientation and Anxiety Symptoms among African American Adolescents," *Journal of Abnormal Child Psychology* (34:3), pp 423-431.

Pang, B., and Lee, L. 2008. "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval* (1:1-2), pp 1-135.

Pang, B., Lee, L., and Vaithyanathain, S. 2002. "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," *the ACL-02 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Philadelphia: ACL Press, pp. 79-86.

Peng, F., Schuurmans, D., Keselj, V., and Wang, S. 2003. "Automated Authorship Attribution with Character Level Language Models," *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, Budapest, Hungary.

Pew Internet and American Life Project. 2008. http://www.pewinternet.org/trends/User_Demo_7.22.08.htm).

Plant, E.A., Hyde, J.S., Keltner, D., and Devine, P.G. 2000. "The Gender Stereotyping of Emotions," *Psychology of Women Quarterly* (24), pp 81-92.

Platt, J. 1999. *Fast Training on Svms Using Sequential Minimal Optimization*, (In Scholkopf, B., Burges, C., and Smola, A. (Ed.) ed.). Cambridge, MA: MIT Press.

Quinlan, J.R. 1986. "Induction of Decision Trees," *Machine Learning* (1:1), pp 81-106.

Riloff, E., and Wiebe, J. 2003. "Learning Extraction Patterns for Subjective Expressions," *the Conference on Empirical Methods in Natural Language Processing*, Morristown, NJ: ACL Press, pp. 105-112.

Rime, B., Mesquita, B., Philippot, P., and Boca, S. 1991. "Beyond the Emotional Event: Six Studies on the Social Sharing of Emotion," *Cognition and Emotion* (5:5-6), pp 435-465.

Sallis, J.F. 2000. "Age-Related Decline in Physical Activity: A Synthesis of Human and Animal Studies," *Medicine and Science in Sports and Exercise* (32:9), pp 1598-1600.

Sallnas, E.L., Rassmus-Grohn, K., and Sjostrom, C. 2000. "Supporting Presence in Collaborative Environments by Haptic Force Feedback," *ACM Transactions on Computer-Human Interaction* (7:4), pp 461-476.

Schapire, R.E., and Singer, Y. 2000. "Boostexter: A Boosting-Based System for Text Categorization," *Machine Learning* (39:2-3), pp 135-168.

Scherer, K.R., Wallbott, H.G., and Summerfield, A.B. 1986. *Experiencing Emotion: A Cross-Cultural Study*. Cambridge, England: Cambridge University Press.

Schler, J., Koppel, M., Argamon, S., and Pennebaker, J. 2006. "Effects of Age and Gender on Blogging," *Proceedings of AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*, Menlo Park, California, pp. 199-205.

ScienceDaily. 2009. "Lifelong Gender Difference in Physical Activity Revealed," in: *ScienceDaily*.

Scott, S., and Matwin, S. 1999. "Feature Engineering for Text Classification," *Proceedings of ICML-99, 16th International Conference on Machine Learning*, pp. 379-388.

Seale, C., Ziebland, S., and Charteris-Black, J. 2006. "Gender, Cancer Experience and Internet Use: A Comparative Keyword Analysis of Interviews and Online Cancer Support Groups," *Social Science and Medicine* (62:10), pp 2577-2590.

Sears, H.A., Graham, J., and Campbell, A. 2009. "Adolescent Boys' Intentions of Seeking Help from Male Friends and Female Friends," *Journal of Applied Developmental Psychology* (30:6), pp 738-728.

Seiffert, C., Khoshgoftaar, T.M., Hulse, J.V., and Napolitano, A. 2010. "Rusboost: A Hybrid Approach to Alleviating Class Imbalance," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* (40:1), pp 185-197.

Shannon, C.E. 1948. "A Mathematical Theory of Communication," *Bell System Technical Journal* (27:4), pp 379-423.

Sherman, A.P. 2001. *Cybergrrl @ Work: Tips and Inspiration for the Professional You*. Berkley Trade.

Shields, S., Garner, D., Di-Leoni, B., and Hadley, A. 2007. "Gender and Emotion," in: *Handbook of the Sociology of Emotions*, J.E. Stets and J.H. Turner (eds.). New York: Springer.

Shim, K.J., Pathak, N., Ahmad, M.A., DeLong, C., Borbora, Z., Mahapatra, A., and Srivastava, J. 2011. "Analyzing Human Behavior from Multiplayer Online Game Logs: A Knowledge Discovery Approach," *IEEE Intelligent Systems* (26:1), pp 85-89.

Short, J., Williams, E., and Christie, B. 1976. *The Social Psychology of Telecommunications*. London, England: John Wiley.

Shye, D., Mullooly, J.P., Freeborn, D.K., and Pope, C.R. 1995. "Gender Differences in the Relationship between Social Network Support and Mortality: A Longitudinal Study of an Elderly Cohort," *Social Science medicine* (41:7), pp 935-947.

Solberg, S., Choi, K.H., Ritsma, S., and Jolly, A. 1994. "Asian American College Students: It's Time to Reach Out," *Journal of College Student Personnel* (35:4), pp 296-301.

Sproull, L., and Kiesler, S. 1986. "Reducing Social Context Cues: Electronic Mail in Organisational Communication," *Management Science* (32), pp 1492-1512.

Stoppard, J.M., and Gunn-Gruchy, C.D. 1993. "Gender, Context, and Expression of Positive Emotion," *Personality and Social Psychology Bulletin* (19:2), pp 143-150.

Subasic, P., and Huettner, A. 2001. "Affect Analysis of Text Using Fuzzy Semantic Typing," *IEEE Transactions on Fuzzy Systems* (9:4), pp 483-496.

Tannen, D. 1991. *You Just Don't Understand: Women and Men in Conversation*. London: Virago Press.

Thelwall, M., Wilkinson, D., and Uppal, S. 2010. "Data Mining Emotion in Social Network Communication: Gender Differences in Myspace," *Journal of the American Society for Information Science and Technology (JASIST)* (61:6), pp 190-199.

Thompson, C. 2001. *Conservation of Resources Theory, a Sloan Work and Family Encyclopedia Entry*. Chestnut Hill, MA: Boston College.

Tonge, J. 2010. "The Influence of Position and Gender on Personal Networks in a Uk Professional Service," *Industrial Marketing Management* (39:3), pp 390-399.

Trost, S.G., Pate, R.R., Sallis, J.F., Freedson, P.S., Taylor, W.C., Dowda, M., and Sirard, J. 2002. "Age and Gender Differences in Objectively Measured Physical Activity in Youth," *Medicine and Science in Sports and Exercise* (34:2), pp 350-355.

Turney, P.D. 2002. "Thumbs up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," *Proceedings of the 40th Annual Meetings of the Association for Computational Linguistics*, Philadelphia, Pennsylvania: ACL Press, pp. 417-424.

Turney, P.D., and Littman, M.L. 2003. "Measuring Praise and Criticism: Inference of Semantic Orientation from Association," *ACM Transactions on Information Systems* (21:4), pp 315-346.

Tweedie, F.J., and Baayen, R.H. 1998. "How Variable May a Constant Be? Measures of Lexical Richness in Perspective.," *Computers and the Humanities* (32:5), pp 323-352.

Varvello, M., and Voelker, G.M. 2010. "Second Life: A Social Network of Humans and Bots," *The 20th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, Amsterdam, The Netherlands: ACM, pp. 9-14.

Walther, J.B. 1992. "Interpersonal Effect in Computer-Mediated Interaction: A Relational Perspective," *Communication Research* (19), pp 50-88.

Wang, F.-Y., Zeng, D., Carley, K.M., and Mao, W. 2007. "Social Computing: From Social Informatics to Social Intelligence," *IEEE Intelligent Systems* (March/April), pp 79-83.

White, S., and Smyth, P. 2003. "Algorithms for Estimating Relative Importance in Networks," *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, Washington D.C., USA: ACM, pp. 266-275.

Wiebe, J., and Riloff, E. 2005. "Creating Subjective and Objective Sentence Classifiers from Unannotated Texts," *the Sixth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, Mexico.

Wiebe, J., Wilson, T., and Bell, M. 2001. "Identifying Collocations for Recognizing Opinions," *Proceedings of the ACL/EACL Workshop on Collocation*, Toulouse, France: ACL Press.

Wiebe, J., Wilson, T., Bruce, R., Bell, M., and Martin, M. 2004. "Learning Subjective Language," *Computational Linguistics* (30:3), pp 277-308.

Wilks, Y., and Stevenson, M. 1998. "The Grammar of Sense: Using Part-of-Speech Tags as a First Step in Semantic Disambiguation," *Journal of Natural Language Engineering* (4), pp 135-144.

Williams, D., Consalvo, M., Caplan, S., and Yee, N. 2009. "Looking for Gender: Gender Roles and Behaviors among Online Gamers," *Journal of Communication* (59:4), pp 700-725.

Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., and Patwardhan, S. 2005a. "Opinionfinder: A System for Subjectivity Analysis," *the Human Language Technology Conference*, Vancouver, Canada.

Wilson, T., Wiebe, J., and Hoffmann, P. 2005b. "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," *the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, Vancouver, Canada.

Witten, I.H., and Frank, E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques (2nd Edition)*, (2nd Edition ed.). San Francisco: Morgan Kaufmann.

Wood, W., and Eagly, A.H. 2002. "A Cross-Cultural Analysis of the Behavior of Women and Men: Implications for the Origin of Sex Differences," *Psychological Bulletin* (128), pp 699-727.

Wood, W., Rhodes, N., and Whelan, M. 1989. "Sex Differences in Positive Well-Being: A Consideration of Emotional Style and Marital Status," *Psychological Bulletin* (106), pp 249-264.

Yang, Y., and Pedersen, J.O. 1997. "A Comparative Study on Feature Selection in Text Categorization," *Proceedings of the ICML97*, pp. 412-420.

Yee, N., and Bailenson, J.N. 2007. "The Proteus Effect: The Effect of Transformed Self-Representation on Behavior," *Human Communication Research* (33), pp 271-290.

Yee, N., and Bailenson, J.N. 2008. "A Method for Longitudinal Behavioral Data Collection in Second Life," *Presence: Teleoperators and Virtual Environments* (17:6), pp 594-596.

Yee, N., Bailenson, J.N., Urbanek, M., Chang, F., and Merget, D. 2007. "The Unbearable Likeness of Being Digital: The Persistence of Nonverbal Social Norms in Online Virtual Environments," *Journal of Cyberpsychology and Behavior* (10), pp 115-121.

Yeh, C.J. 2002. "Taiwanese Students' Gender, Age, Interdependent and Independent Self-Construal, and Collective Self-Esteem as Predictors of Professional Psychological

Help-Seeking Attitudes," *Cultural Diversity and Ethnic Minority Psychology* (8:1), pp 19-29.

Youngs, G. 2004. "Cyberspace: The New Feminist Frontier," in: *Women and Media: International Perspectives* K. Ross and C.M. Byerly (eds.). Wiley-Blackwell, pp. 185-208.

Yule, G.U. 1944. *The Statistical Study of Literary Vocabulary*. Cambridge University Press.

Zhang, Y., Yu, X., Dang, Y., and Chen, H. 2010. "An Integrated Framework for Collecting and Analyzing Avatar Data from Virtual World: An Exploratory Study in Second Life," *IEEE Intelligent Systems* (25:6), pp 17-23.

Zheng, R., Li, J., Chen, H., and Huang, Z. 2006. "A Framework for Authorship Identification of Online Messages: Writing-Style Features and Classification Techniques," *Journal of the American Society for Information Science and Technology (JASIST)* (57:3), pp 378-393.

Zhou, L., and Chaovalit, P. 2008. "Ontology-Supported Polarity Mining," *Journal of the American Society for Information Science and Technology (JASIST)* (59:1), pp 98-110.