



Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2015

A WEB PERSONALIZATION ARTIFACT FOR UTILITY-SENSITIVE REVIEW ANALYSIS

Long Flory Mrs.

VIRGINIA COMMONWEALTH UNIVERSITY, floryl@vcu.edu

Follow this and additional works at: <http://scholarscompass.vcu.edu/etd>

 Part of the [Business Intelligence Commons](#), [E-Commerce Commons](#), [Marketing Commons](#), and the [Technology and Innovation Commons](#)

© The Author

Downloaded from

<http://scholarscompass.vcu.edu/etd/3739>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

©Long Flory 2015
All Rights Reserved

**A WEB PERSONALIZATION ARTIFACT
FOR UTILITY-SENSITIVE REVIEW ANALYSIS**

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Business at Virginia Commonwealth University

By

Long Flory

Master of Science in Decision Science & Business Analytics
Virginia Commonwealth University
December, 2009

Director: Dr. Kweku-Muata Osei-Bryson
Professor, Information Systems

Virginia Commonwealth University
Richmond, Virginia
April, 2015

ACKNOWLEDGEMENT

I would like to take this moment to express my deepest gratitude to all those who gave their advices, time, and resources important to the accomplishment of this dissertation. I thank my committee members for all advices they have given me and all things they have done for me. My special thanks go to my Advisor, Dr. Osei-Bryson, for his support and insight. I very appreciate the detailed considerations that I received from him. I thank Dr. Dula for his stimulating questions that helped me refine my design ideas. Dr. Weistroffer provided valuable advice on the presentation of this work. Dr. Thomas' challenging questions helped me improve the quality of the research solution. I thank the faculty and staff of the Information Systems Department and the Graduate Office of Business School of Virginia Commonwealth University. Their assistance was valuable for the completion of this research. I thank my friends: Basle, Brain, Bill, Christy, David, Edward, James, Janie, Julie, Lian, Linda, Mark, Shu, Tammy, Tian, Yang, Yu, and Walter. They provided necessary help for me to improve the rigor and the relevance of this research. I am very grateful to my husband for his understanding, sacrifice, encouragement, and support. I appreciate my family's understanding and help during the time of this research. I cherish you all!!

TABLE OF CONTENTS

LIST OF FIGURES.....	vi
LIST OF TABLES.....	vii
ABSTRACT.....	viii
CHAPTER 1 INTRODUCTION.....	1
1.1 Background.....	5
1.1.1 Online Review.....	5
1.1.2 Review Helpfulness.....	7
1.1.3 Review Helpfulness Research.....	8
1.1.4 The Important Roles Of Review Helpfulness Research.....	9
1.2 Limitations Of Prior Review Helpfulness Research.....	11
1.3 Research Problem and Objectives.....	13
1.4 Research Contributions.....	14
1.5 Dissertation Organization.....	15
CHAPTER 2 REVIEW OF THE LITERATURE.....	16
2.1 Web-based Consumer Decision Support Research.....	16
2.1.1 Product Webpage Effectiveness.....	17
2.1.2 Personalized Consumer Decision Support.....	19
2.2 Web Personalization Research.....	21
2.2.1 Input Processing.....	22
2.2.2 Webpage Searching.....	25
2.3 Review Helpfulness Research.....	30
2.3.1 Supervised Techniques with Helpful Ranks.....	31
2.3.2 Supervised Techniques without Helpful Ranks.....	35
2.3.3 Unsupervised Techniques.....	37
2.3.4 Personalized Review Helpfulness.....	38

CHAPTER 3 RESEATCH METHODOLOGY.....	40
3.1 Behavioral Science and Design Science.....	41
3.2 Influential Literature of Design Science Methodology	43
3.3 Our Research Methodology	48
CHAPTER 4 MODEL AND ARCHITECTURE.....	54
4.1 USRAnalyzer Overview Model	55
4.2 A Use Case of USRAnalyzer.....	57
4.3 USRAnalyzer Architecture.....	65
4.3.1 User Search Initiation.....	66
4.3.2 Input Processing.....	67
4.3.3 Relevance Evaluation.....	69
4.3.4 User Relevance Criterion, Review Selection, Helpful Review Presentation.....	71
CHAPTER 5 ALGORITHMS.....	73
5.1 Directed Particle Swarm Optimization and K-Means Algorithm.....	73
5.1.1 Prior PSO-KM Method.....	73
5.1.2 Proposed DPSSO-KM Algorithm.....	76
5.2 Review Utility Ranking Algorithm.....	86
CHAPTER 6 IMPLEMENTATION.....	90
6.1 User Interface Implementation	90
6.2 Back End Implementation	91
CHAPTER 7 EXPERIMENTAL EVELUATIONS	94
7.1 Overview of Experiments.....	94
7.2 Evaluation Metrics.....	96
7.3 Experiment I.....	97
7.3.1 Experiment I Data Collection.....	98
7.3.2 Experiment I Design.....	99
7.3.3 Experiment I Outcomes	104
7.4 Experiment II.....	105

7.4.1 Experiment II Data Collection	106
7.4.2 Experiment II Design.....	106
7.4.3 Experiment II Outcomes	108
7.5 Implications of Experiments.....	110
CHAPTER 8 FUTURE RESEARCHES.....	114
CHAPTER 9 CONCLUSIONS.....	118
LIST OF REFERENCES.....	122
APPENDIX I	
K-Means Clustering Overview.....	149
APPENDIX II	
Support Vector Machine Overview.....	152
APPENDIX III	
Decision Tree Overview.....	156

LIST OF FIGURES

Figure 1. An Example of Online Review.....	2
Figure 2. Design Science Framework by Hevner et al. (2004).....	47
Figure 3. Design Evaluation Method Suggested by Hevner et al. (2004).....	51
Figure 4. USRAnalyzer Model.....	57
Figure 5. A Large Review Pool at tripadvisor.com.....	59
Figure 6. An Example of Consumer Inputs.....	60
Figure 7. An Example of Feedback to Consumer.....	61
Figure 8. An Example of Desirable Relevance Level.....	62
Figure 9. An Example of Resultant Helpful Reviews.....	64
Figure 10. USRAnalyzer Architecture.....	66
Figure 11. G-means from Hamerly & Elkan (2004).....	77
Figure 12. DPSO-KM Pseudo Code.....	80
Figure 13. Example II, Two First-Cycle Clusters.....	83
Figure 14. Example II, Final KM Clusters.....	84
Figure 15. Example II, Relevant Set (RS).....	85
Figure 16. RURanking Algorithm.....	89
Figure 17. An Example of Star Rating at tripadvisor.com.....	99
Figure 18. HCEB's consumer needs.....	100
Figure 19. Examples of Consumer Inputs.....	101
Figure 20. K-Means Results from Figure 4 of Jain (2010).....	150
Figure 21. Support Vector Machine from Figure 1 of Marron (2015).....	154
Figure 22. AdaBoost.....	158

LIST OF TABLES

Table 1. Literature Review Summary (I).....	21
Table 2. Literature Review Summary (II).....	30
Table 3. Literature Review Summary (III).....	39
Table 4. Experiment I Tuning Experiements.....	103
Table 5. Experiment I Outcomes.....	105
Table 6. Experiment II Tuning Experiements.....	108
Table 7. Experiment II Outcomes.....	110

ABSTRACT

A WEB PERSONALIZATION ARTIFACT FOR UTILITY-SENSITIVE REVIEW ANALYSIS

By Long Flory, Ph.D.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Business at Virginia Commonwealth University

Virginia Commonwealth University, 2015.

Major Director: Dr. Kweku-Muata Osei-Bryson, Professor, Information Systems

Online customer reviews are web content voluntarily posted by the users of a product (e.g. camera) or service (e.g. hotel) to express their opinions about the product or service. Online reviews are important resources for businesses and consumers. This dissertation focuses on the important consumer concern of review utility, i.e., the helpfulness or usefulness of online reviews to inform consumer purchase decisions. Review utility concerns consumers since not all online reviews are useful or helpful. And, the quantity of the online reviews of a product/service tends to be very large. Manual assessment of review utility is not only time consuming but also information overloading. To address this issue, review helpfulness research (RHR) has become a very active research stream dedicated to study utility-sensitive review analysis (USRA) techniques for automating review utility assessment.

Unfortunately, prior RHR solution is inadequate. RHR researchers call for more suitable USRA approaches. Our current research responds to this urgent call by addressing the research problem: What is an adequate USRA approach? We address this problem by offering novel Design Science (DS) artifacts for personalized USRA (PUSRA). Our proposed solution extends not only RHR research but also web personalization research (WPR), which studies web-based solutions for personalized web provision. We have evaluated the proposed solution by applying three evaluation methods: analytical, descriptive, and experimental. The evaluations corroborate the practical efficacy of our proposed solution.

This research contributes what we believe (1) the first DS artifacts to the knowledge body of RHR and WPR, and (2) the first PUSRA contribution to USRA practice. Moreover, we consider our evaluations of the proposed solution the first comprehensive assessment of USRA solutions. In addition, this research contributes to the advancement of decision support research and practice. The proposed solution is a web-based decision support artifact with the capability to substantially improve accurate personalized webpage provision. Also, website designers can apply our research solution to transform their works fundamentally. Such transformation can add substantial value to businesses.

KEYWORDS: online review, review helpfulness, utility-sensitive review analysis, text mining, data mining, personal and dynamic review utility analysis, web personalization, web 2.0 technology, web-based decision support, consumer decision support

CHAPTER 1 INTRODUCTION

Online customer reviews, known as online reviews or reviews, are web content at eCommerce websites (e.g. amazon.com) and other websites (yooqa.com). Online reviews are voluntarily posted by the users of a product (e.g. camera, bicycle) or service (e.g. hotel, restaurant) to express their opinions about the product or service. Figure 1 is a snapshot of the scanner CanoScan LiDE110's reviews at amazon.com. The review in Figure 1 receives a helpful ratio, 520 of 527, i.e. "520 of 527 people found the following review helpful." The helpful ratio indicates that 527 readers of the review gave it a utility vote that answers the question, "Was this review helpful to you?" And, 520 out of the 527 readers gave a yes (helpful) vote. The helpful ratio determines the helpful rank of the review, which is the top most or most helpful. Increasingly, more and more consumers use online reviews in their shopping. Figure 1 shows that total 1939 readers have given a vote to the two top-ranked reviews.

Like the reviews at amazon.com, online reviews are valuable to businesses and consumers. Businesses value online reviews as a business intelligence (BI) resource. Such resource enables businesses to understand consumer purchase decision (Gilbert & Karahalios, 2010). Many firms consider online reviews valuable for business competitiveness (Chung et al., 2005; Marshall et al., 2004). Meanwhile, a large number of consumers use online reviews to make better purchase decisions. In general, consumers regard online reviews as trustworthy resources (Hopkins & Sarner, 2013; Montoyo et al., 2012).

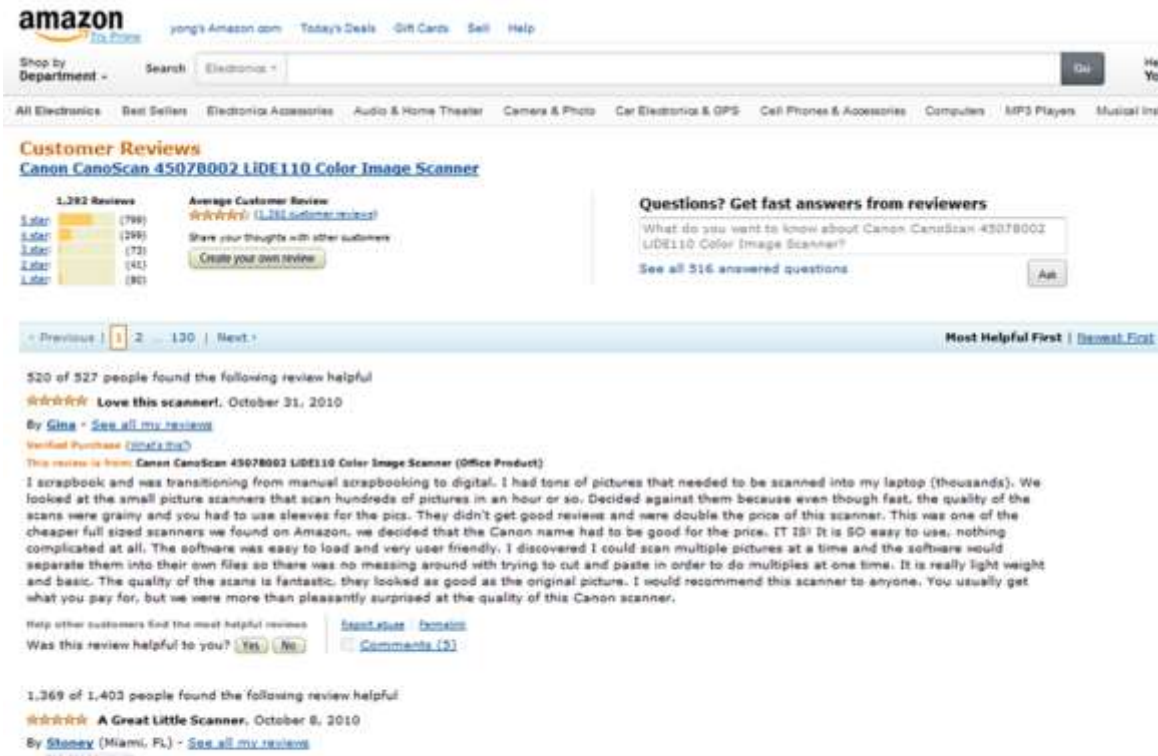


Figure 1. An Example of Online Review

For consumers, a central concern is review utility, i.e., the helpfulness or usefulness of a review to inform consumers' purchase decisions. This concern is increasingly important as it is observed the exponential growth of online reviews (Yin et al., 2014). A product or service often receives thousands reviews on a review website. For instance, at the time of our visit of amazon.com, CanoScan LiDE110 had received 1369 reviews on 135 webpages. The large quantity of online reviews has caused the problem of information overload. Especially, not all the reviews are helpful or useful to consumers who typically want to find the useful reviews (Liu

et al., 2012). However, manual selection of useful reviews is intractable to consumers when the review pool is large.

In order to help consumers reduce information overload, review websites usually include helpful ranking to list the most helpful reviews on the top. So, consumers can read the top-ranked reviews. But, there is a serious problem with the helpful ranking, which researchers refer to as missing vote (Liu et al., 2007); that is, many reviews do not receive enough votes because a large number of review readers do not vote. Those reviews are placed behind the top-ranked helpful reviews, yet potentially useful to consumers' purchase decisions. It can be seen that low-ranked reviews are not necessarily unhelpful (Liu et al., 2007) but the helpful ranking list suppresses their utilities. This is an inherent problem of the helpful ranking, which needs to be addressed (Chen & Tseng, 2011; Liu et al., 2007).

As an active research stream, review helpfulness research (RHR) is aimed at proposing automatic approaches of utility-sensitive review analysis (USRA) that is the automatic process to assess the utility of an online review. Prior RHR has proposed various models that can predict the helpfulness or utility of every review on review websites. Unfortunately, the existing models have significant weaknesses that impose a number of challenges difficult to solve. To address them, RHR has become a hot research stream in multiple disciplines such as information systems (IS), computer sciences, and computational linguistics. A major weakness in prior RHR is that most RHR models do not provide personalized USRA (PUSRA) (Mudambi & Schuff, 2010; Yin et al., 2014). PUSRA approach is needed since individual consumers have different purchase needs. A review meeting the need of one person may not meet the needs of the others. A review

is helpful to one consumer may be unhelpful to the others (Moghaddam et al., 2012). For example, the review “the cellphone has a poor camera” is useful to the consumers who need good cellphone camera. But, that review may not helpful to those who need easy typing screen and do not regard a good cellphone camera as an important concern.

A few researches provide personalized models for PUSRA. Still, they cannot adequately address the personal need of consumers since prior personalized solution is still predictive models with significant limitation (Yin et al., 2014). For instance, when a consumer choose to read the helpful reviews predicted by a personalized model, it then uses that consumer’s choices to predict the helpful reviews for the next person with similar demographic characteristics as that of the first consumer. Such prediction can be a mistake since the two persons could have quite different needs although they have similar demographic characteristics. So, the reviews useful to the first consumer may be not helpful to the second consumer. Therefore, prior personalized solution needs to be improved in order to provide more adequate PUSRA. And, effective PUSRA approach calls for a web personalization (WP) artifact, which is a web-based information system offering individual consumers with personalized helpful reviews better meeting their needs. Currently, such a WP artifact is lacked in the literature. We attempt to fill out this research gap by addressing the research problem: What is an effective WP artifact for PUSRA?

To address the research problem, we propose the Interactive Utility-Sensitive Review Analyzer (USRAnalyzer), which is an interactive WP artifact. It enables the interaction between the USRAnalyzer and the consumer to meet personal need. Our solution is what we believe, the

first DS contribution for PUSRA, which is grounded on IS design science principle. Thus, this research is an information systems research rather than a computer-science or linguistic one. The goal of this research is to (1) contribute substantial new knowledge to DS research in general, to RHR and WPR in particular; (2) maximize the value of online reviews to IS researchers, IS professionals, consumers and businesses.

The structure of this INTRODUCT is as followed. Section 1.1 provides background knowledge of online review, review helpfulness, and review helpfulness research (RHR). Section 1.2 discusses the limitations of prior RHR. Section 1.3 presents the research problem addressed by this research. Section 1.4 outlines the contributions of our current research. Section 1.5 outlines the structure of this dissertation.

1.1 Background

This section provides the background knowledge of online reviews, review helpfulness, and review helpfulness research (RHR). We discuss online reviews in Section 1.1.1. We introduce review helpfulness in Section 1.1.2 and review helpfulness research (RHR) in Section 1.1.3. Section 1.1.4 presents the important roles of review helpfulness research.

1.1.1 Online Review

In IS literature, an online review refers to a product/service review given by a website user known as reviewer who expresses his/her opinion on the product/service (Montoyo et al., 2012; Pang et al., 2002; Terveen et al., 1997). The reviewer is generally a consumer who has used the product/service. A professional reviewer may evaluate the product/service and post

his/her evaluation on the review website. Review helpfulness research (RHR) focus on the reviews posted by the product/service users (Chen & Xie, 2008; Mudambi & Schuff, 2010). So, our research is concerned about product- or service-users' reviews rather than professionals' evaluations.

Online reviews facilitate consumers who search from the opinions of the others about certain products/services. And, online reviews have improved consumers' purchase decisions (Dellarocas, 2003; Duan et al., 2008; Woodcock et al., 2011). Recently, online reviews become primary opinion sources for most consumers who believe online reviews are credible and trustworthy (Montoyo et al., 2012; Lu et al., 2014). More and more consumers utilize online reviews in their purchase decisions (Goh et al., 2013). Since online reviews influence consumers' purchase decisions, firms use online reviews as business assets for competitiveness (Dellarocas, 2003; Gilbert & Karahalios, 2010; Li & Hitt, 2008). The uses of online reviews have led to better business decisions (Narver & Slater, 1990; Park & Kim, 2009). Researches have reported significant increase in sales revenue from using online reviews in business intelligence (Gilbert & Karahalios, 2010; Liu et al., 2014). The eCommerces with a large number of online reviews have experienced substantial business growth (Archak et al., 2011; Chevalier & Mayzlin, 2006; Korfiatis et al., 2012).

However, there is a very challenging issue related to the utility of online reviews to consumer purchase decisions. This issue is known as review helpfulness or review utility (Chen et al., 2008; Korfiatis et al., 2012; Wolfinbarger & Gilly, 2001), which we discuss in section 1.1.2.

1.1.2 Review Helpfulness

Review helpfulness refers to the usefulness or utility of a review to consumers' purchase decisions. In review helpfulness research (RHR), these three terms are synonyms: review helpfulness, review usefulness, and review utility (Ghose & Ipeiritis, 2007; Moghaddam et al., 2010; Montoyo et al., 2012). Prior RHR involves two distinct assumptions about review utility, which are common utility assumption and personal utility assumption. The former assumes if a review is helpful to one consumer, it is useful to the other. The latter assumes the helpfulness of a review differs from one consumer to another; a review helpful to one person may be not helpful to the other. Most RHR assumes common utility (e.g. Ghose & Ipeiritis, 2011; Mudambi & Schuff, 2010; Ngo-Ye & Sinha, 2012). So far, only a few RHR works promote personal utility (e.g. Moghaddam et al., 2010; Montoyo et al., 2012).

In this paper, we assume personal utility since consumers usually have different purchase needs. A review meeting one need may not satisfy a different need. Thus, a review is not commonly useful to everybody (DeBono et al., 2003; Montoyo et al., 2012). For example, suppose a consumer wants to buy a car and needs high engine power. The follow review is likely helpful to that consumer: "Model xx is very powerful, accelerating faster than its peer." Distinctly, another consumer also wants to buy the same car but needs high gas mileage. To the second consumer, the previous review is probability not helpful.

1.1.3 Review Helpfulness Research

Review helpfulness research (RHR) studies effective solutions for utility-sensitive review analysis (USRA), i.e., the automatic process to identify helpful reviews for consumers. The essential goal of RHR is to reduce information overload to consumers who are concerned to review utility (Cao et al., 2011; Gilbert & Karahalios, 2010; Nassirtoussi et al., 2014). Also, RHR addresses the inherent problem of the helpful ranking used by review websites, which positions the most helpful reviews at the top according to the votes of review readers. Since many readers do not vote, a large number of online reviews receive none or few vote(s). They receive inadequate number of votes, and thus, very low helpful ranks behind tens or hundreds helpful reviews. For example, Liu et al. (2007) find that over 70% of the reviews at amazon.com received few or none votes. Cao et al. (2011) reported 51% of the reviews on CNETD received no vote. Researchers generally agree that the problem in the helpful ranking is mainly caused by the manual voting approach by which review readers may not vote.

The inherent problem in the helpful ranking suggests that the reviews receiving few or none votes are not necessarily unhelpful. Low helpful ranks tend to skew the true values of the corresponding reviews. Liu et al. (2007) compared manually-constructed helpful ranks with helpful voting ranks. They found substantially gap between the two. Like Liu et al. (2007), other researches have also concluded that helpful ranking approach suppresses useful reviews (Pang & Lee, 2008; Ngo-Ye & Sinha, 2014). Review helpfulness research (RHR) addresses the problem with automatic solutions for utility-sensitive review analysis (USRA) (Ghose & Ipeiritis, 2011; Nassirtoussi et al., 2014; Ngo-Ye & Sinha, 2014).

In general, RHR includes the study of automatic USRA approaches. However, a few researches include review spam detection in RHR (e.g. Cao et al., 2011; Wang et al., 2011). Review spam detection studies the fake reviews (review spams) posted by some individual to promote or discredit a product/service or a firm. Review spams usually contain false information harmful to consumers. USRA approaches should not use review spams. However, most RHR researchers regard review spam detection should be a separate research stream apart from RHR (Pang & Lee, 2008). The reason for the majority's view relates to the focus of review spam detection research. Although review spams threaten review helpfulness analysis they are more a complex behavior problem outside review utility concern (Lim et al., 2010; Ott et al., 2011). Thus, most RHR assumes online reviews are free from review spams (Cheung & Lee, 2012; Piramuthu et al., 2012). In this paper, we agree with most RHR researchers. We consider review spam detection is beyond the scope of our research. We assume that online reviews are free from review spams.

1.1.4 The Important Roles of Review Helpfulness Research

To date, review helpfulness research (RHR) has played an important role in both practices and researches associated with online reviews. RHR becomes more and more important in meeting the needs of IS designers, IS researchers, consumers, and businesses.

First of all, RHR helps the designers of review websites replace helpful ranking with USRA, which can help the businesses improve the understanding of their consumers (Liu, 2012; Montoyo et al., 2012; Pang & Lee, 2008). In section 1.1.3, we discuss that the helpful ranking

approach can inaccurately present the utility of an online review. With inaccurate helpful votes, it is difficult for businesses to understand their customers.

Secondly, RHR contributes effective solutions USRA that is aimed to eliminate the problems of helpful ranking and information overload simultaneously (Chen & Tseng, 2011; Liu, 2010; Pang & Lee, 2008). USRA approach can help consumers identify helpful reviews with higher accuracy and minimum information overload comparing to the helpful ranking approach (Chen & Tseng, 2011; Liu et al., 2007; Ngo-Ye & Sinha, 2012). As such, may researchers call for increasing RHR effort (Danescu-Niculescu-Mizil et al., 2009; Moghaddam et al., 2012). Recently, we have seen more published studies in RHR (Montoyo et al., 2012; Liu et al., 2013).

Thirdly, RHR helps eCommerce vendors achieve their business objective. By implementing effective USRA, the eCommerces can improve the accuracy of helpful review provisions (Clemons et al., 2006). eCommerce vendors now rely on helpful ranking. However, the resultant helpful reviews are not satisfactory to consumers (Cao et al., 2011; Liu et al., 2012). RHR is critical to help eCommerce vendors implement better solutions, and empower the businesses with the capability to meet customers' needs. This capability has enabled eCommerces to draw enormous number of consumers and generate substantial business growth (Danescu-Niculescu-Mizil et al., 2009; Otterbacher, 2009). The business growth is expected to be more significant when effective personalized utility-sensitive review analysis (PUSRA) is implemented (Cao et al., 2011).

1.2 Limitations of Prior Review Helpfulness Research

Along with growing importance of RHR, the researchers have recognized the major limitations of the existing research contributions. This section presents the limitations. Particularly, we show the limitation of common utility assumption and the limitation of prior personalized USRA solution.

The readers may recall that RHR may assume common utility or personal utility. Common utility assumption assumes the helpfulness of a review is the same to all consumers. Most prior research is based on common utility. Under it, the researchers have proposed various models to predict review utility. Alternatively, personal utility assumption assumes the utility of a review is different to different consumers.

Recently, some prominent researchers have exposed the serious limitation of the common utility assumption. They argue that consumers do not commonly agree on the utility of a review. In other words, a review may have different utilities to two consumers because of their needs may not be the same (Rilof et al., 2006). Montoyo et al. (2012) call for effective approaches to provide personalized utility-sensitive review analysis (PUSRA). They suggest that the analysis should focus on individual needs, replacing common utility assumption with personal utility assumption. The researchers suggested that most prior RHR suffered the limitation of common utility assumption.

Assuming personal utility, several works (e.g. Montoyo et al., 2012, Cao et al., 2011) proposed the algorithms in the attempt of personalizing review utility. Still, the solutions are

limited since they did not provide a web personalization (WP) artifact, the information system automating the provision of personalized reviews. Particularly, the existing RHR solutions lack an interactive WP artifact for suitable personalized provision of helpful reviews. The severity of this RHR limitation is shown in the next paragraph.

WP includes two types, record-based personalization and interactive personalization (Johar et al., 2014; Wang & Benbasat, 2014). The former provides personalized reviews according to the records of users' past visits. The latter provides personalized reviews according to users' runtime activities. Typically, interactive personalization can better meet consumer needs than can record-based personalization (Bhargava et al., 2007; Dong et al., 2004; Johar et al., 2014). This is because the latter is limited to user models derived from users' past searches. The user models tend to limit the user need to the past. When a consumer searches for online reviews to make his/her purchase decision, the need of the consumer is influenced by the reviews presented to him/her. The need of the consumer may change during his/her interaction with the review website (Agarwal & Karahanna, 2000; Burton-Jones & Grange, 2013; McKinney & Yoos, 2010). The consumer's need changes during the review search (Eroglu et al., 2003; Parboteeah et al., 2009). Record-based personalization cannot capture the changes while the runtime-based web personalization can address them via the interaction between the WP artifact and the user (Pavlou et al., 2007; Wang & Benbasat, 2009). Therefore, effective RHR solutions need an interactive WP artifact. Otherwise, the utility of online reviews can be misrepresented given the changes of consumer need.

So far, no RHR addresses the limitation of prior personalized USRA (PUSRA) approach. There is not an interactive WP artifact for PUSRA in the literature. Our current research is attempted to close this literature gap by answering the research problem and objectives presented in section 1.3.

1.3 Research Problem and Objectives

As addressed in Section 1.2, utility-sensitive review analysis (USRA) needs an interactive web personalization (IWP) artifact. In the absence of such system in the literature, we attempt to fill the gap by addressing the research problem: What is an effective and efficient IWP artifact that can provide personalized utility-sensitive review analysis (PUSRA) meeting the changing needs of individual consumers? The issues we seek to address in this research are listed as following:

1. What are suitable overview models for the PUSRA artifact?
2. What are suitable architectures for the PUSRA artifact?
3. What are suitable methods to identify helpful reviews meeting personal needs?
4. What are suitable evaluations to assess the effectiveness and efficiency of the PUSRA artifact?

We attempt to achieve three research objectives. First, we bring significant improvement into review helpfulness research. Second, we contribute effective methods to web personalization research, which can adopt our proposed methods for significant improvement.

Third, we offer a widely-applicable interactive web personalization solution that the designers of review websites can use to fundamentally improve their designs.

1.4 Research Contributions

Our research has made significant theoretical and practical contributions to Information Systems (IS) discipline as well as to businesses.

Theoretical contributions: Our research answers the question, how to design an effective and efficient PUSRA artifact for personalized utility-sensitive review analysis (PUSRA). Our answer provides the theoretical contribution of design science (DS) that answer how-to-do questions (Hevner et al., 2004; Hevner, 2007).

Contributions to IS Research: This research is a DS research in IS discipline; we contribute novel and useful artifacts (PUSRA artifact and its components) to solve the unsolved IS research problem. Our proposed artifacts address the fundamental limitations of prior review helpfulness research (RHR). We pioneer the PUSRA artifact that will extend the existing RHR solutions and stimulate continuous research effect in PUSRA.

Contributions to IS Practice: The proposed PUSRA artifact and its components inform the designers of review websites about how to significantly improve the website design. Moreover, website designers are looking for new, viable solutions to advance consumer-centric web design (Johar et al., 2014; Wang & Benbasat, 2014). Our solution will provide valuable, practical guidance to advance web-based, consumer-centric systems.

Business contributions: Comparing to prior RHR solution, our proposal can better meet consumers' needs and improve purchase decisions. Doing so is important criterion for consumer acceptance and satisfaction of online reviews (Haubli et al., 2000; Kohli et al., 2004; Xu et al., 2014). Satisfied consumers more likely use review websites. Increased usages support the success of review websites (Eroglu et al., 2003; Parboteeah et al., 2009), which in turn increases the sales and profits of the businesses (Cenfetelli et al., 2008; Jiang et al., 2010; Johar et al., 2014).

1.5 Dissertation Organization

The rest of this dissertation provides the following contents. Chapter 2 reviews the literature in three research streams: (1) web-based consumer decision support research, (2) web personalization research, and (3) review helpfulness research. Chapter 3 discusses the research methodologies in information systems (IS) research and the methodology we use in this research. Chapter 4 presents the overview model and architecture of our proposed PUSRA artifact. Chapter 5 discusses the proposals of two algorithms that play the key roles in PUSRA. Chapter 6 outlines the prototype implementation of the proposed PUSRA artifact. Chapter 7 presents the two experiments by which we evaluation the utility of our solution. Chapter 8 discusses the implementations of the experiment findings. Chapter 9 provides the outlook of future researches. Chapter 10 is the conclusion of this dissertation.

CHAPTER 2 REVIEW OF THE LITERATURE

This chapter reviews three research streams, which are closely related to our research. We review prior research in web-based consumer decision support in section 2.1, prior web personalization research in section 2.2, and prior review helpfulness research in section 2.3.

2.1 Web-based Consumer Decision Support Research

Web-based consumer decision support research attempts to address the factors of effective personalized webpages. Prior research consistently concludes that webpage utility (the helpfulness of a webpage) depends on the needs of individual consumers. There are complex factors influencing webpage utility (Kim & Lennon, 2008; Moe, 2003). For example, product knowledge influences individual's evaluation about the utility of the webpages that describe product price (Lynch & Ariely, 2000). Consumer browsing behavior affects webpage utility (Smyth & Balfe, 2006; van Dijck, 2009). Recent research emphasizes the match between webpage provision and consumer need. Such match is found to be a main factor of webpage utility (Ho & Bodoff, 2014; Li et al., 2014; Smith et al., 2011).

The examples given above reflect the active effort of web-based consumer decision support research. Thousands researches have contributed equivocal findings. And, new findings continue coming up in the literature. It is difficult for us to provide a complete coverage of published findings. For a good coverage of the findings most relevant to our research, this section primarily reviews prominent IS research of web-based consumer decision support. We

review the literature concerning two topics: (1) product webpage effectiveness, i.e., the effect of product webpage on purchase decision (section 2.1.1); (2) personalized consumer decision support, i.e., the impact of personalized web-based consumer decision support on user adoption of the website (section 2.1.2).

2.1.1 Product Webpage Effectiveness

The research on the effect of product webpage commonly supports personal utility assumption. In section 1.2, we discussed personal utility assumption; that is, a webpage useful to one consumer may be unhelpful to the other. In general, prior research agrees that the match between webpage provision and consumer need is the key to high webpage utility (Kuruzovich et al., 2008; Li et al., 2014; Smith et al., 2011). However, Research findings do not agree on other factors of webpage personalization. Here discusses the influential findings contributed by the following researches.

Clemons et al. (2006) studies online reviews on the sales of craft beer. They found that the same review on the Internet did not have equal effect on individual beer consumers. The effect of a review depended on the taste of a specific consumer. The researchers suggested that website providers should tailor online reviews to match individual tastes. Kuruzovich et al. (2008) studied the influence of webpage on consumer web search in automotive retailing industry. They found distinct uses of price and product webpages by different consumers. Also, the utility of a product webpages varies with consumer Internet experience. Kuruzovich et al.

(2008) suggested that businesses should use different web designs in order to offer personalized product webpages.

A large number of researches studied eStore websites (e.g. DeBono et al., 2003; Jiang & Benbasat, 2007; Haeubl & Trifts, 2000; Kohli et al., 2004). Prior research on eStore websites particularly emphasized the individual needs of consumers. Smith et al. (2011) found that unmet consumer need closely related to high product return. And, online consumer support was effective when webpages met individual needs. Similarly, Li et al. (2014) found sale losses when product webpages do not fit in individual needs. A common conclusion was that eStores should update their websites by using web personalization systems. The updating could substantially increase the usages of eStore websites. One-size-for-all webpages do not provide good consumer decision support (Haeubl & Trifts, 2000; Jiang & Benbasat, 2007; Li et al., 2014; Smith et al., 2011).

Some research investigated the factors of consumer attitude towards the use of product webpage. Prior research stresses the accessibility of product webpage. The accessibility was found to be a primary factor of consumer attitude towards the use of product webpage. Accessing to needed webpages led to satisfied use experience (Clemons, 2008; Granados et al., 2012; Li et al., 2014). According to that finding, the researchers emphasized the importance of meeting consumer needs in web-based consumer decision support. Thus, website providers were suggested using web personalization techniques to help individual consumers find the relevant webpages. Without web personalization, product webpages might trigger negative consumer

attitude towards the use of the products (Degeratu et al., 2000; Gu et al., 2012; Kuruzovich, 2008; Lynch & Ariely, 2000).

Additionally, cognitive effect on web-based consumer decision support has been studied. Park & Kim (2008) studied cognitive effect of online reviews. They illustrated that cognitive ability varied from person to person. Review websites should provide the reviews on the basis of individual cognitive ability. And, a review was useful to a consumer when it aligned with the cognitive ability of that consumer. In the same vein, other researchers found that different consumers looked for different online reviews due to the difference in cognitive ability (Bakos, 1997; Gu et al., 2012; Oliver, 1995).

2.1.2 Personalized Consumer Decision Support

The research studying personalized consumer decision support generally agrees with the research studying product webpage effectiveness. However, the former emphasizes cognitive impact of web personalization (Haubli & Trifts, 2000; Ho & Bodoff, 2014; Hong et al., 2004; Kim & Lennon, 2008; Komiak & Benbasat, 2006; Moe, 2003). Several researches offered comprehensive comparison between different personalization approaches and valuable insights on personalized consumer decision support. Wang & Benbasat (2014) compared two approaches: consumer-guided and system-controlled. Consumer-guided approach accepts consumer input with minimum control on input content. System-controlled approach provides a list of choices to consumers who make selection from the list. Wang & Benbasat (2014) found consumer-guided approach could better meet personal need than system-controlled approach in initial interaction

between user and system. The webpages given by consumer-guided approach tend to be more useful to the user than those given by system-controlled approach. Ho & Bodoff (2014) found that purchase motivation and cognitive ability determined the effectiveness of personalized consumer decision support. Johar et al. (2014) compared record-based and interactive web personalization. They concluded that interactive web personalization was a better approach in terms of consumer need and cognitive ability.

More recent researches examine cognitive and emotional outcome of personalized consumer decision support (Parboteeah et al., 2009; Shiv & Fedorikhin, 1999; Song et al., 2008). Lee & Benbasat (2011) and Xu et al. (2014) studied how transparency level of product webpages influence personalized consumer decision support. They argued that the transparency needed to be appropriate in order to satisfy consumers. Since emotion and cognition changes during web-search process, personalized consumer decision support needs to adjust webpages during the interaction between user and system. Lee & Benbasat (2011) provided similar finding. They emphasized the fitness of webpage to user's emotion and cognition. Komiak & Benbasat (2006) studies consumer acceptance of personalized consumer decision support. They found that personalized webpages were trustworthy and useful to a consumer when the webpages were relevant to the need of the consumer (Komiak & Benbasat, 2006).

Furthermore, Parboteeah et al., (2009) concluded that task relevance was the most important requirement for personalized consumer decision support system. Task relevance means the webpages meet consumer needs. Also, mood-relevance influences the effectiveness of personalized consumer decision support. Mood-relevance means the webpages meet personal

taste. Zhang (2000) found the importance of task relevance. The researcher illustrated that inappropriate webpages could distract and frustrate consumers. Parboteeah et al. (2009) and Zhang (2000) agreed that task relevance was dynamic during web search. When a webpage is relevant to the need of a consumer, he/she accepts it. After reading the webpage, the consumer typically changes her/his need or task relevance. Thus, personalize consumer support needs to address the change of task relevance.

In Table 1, we present the main findings of web-based consumer decision support research.

Table 1. Literature Review Summary (I)

Web-based Consumer Decision Support Research		
Topics	Factors	Example Researches
Product Webpage Effectiveness	web personalization	Kuruzovich et al. 2008; Smith et al. 2011
	individual need	Clemons et al. 2006; DeBono et al. 2003; Jiang & Benbasat 2007
	demographic fit	Kuruzovich et al. 2008; Gu et al. 2012
	psychological fit	Park & Kim 2008; Oliver 1995
	webpage characteristics	Granados et al. 2012; Li et al. 2014
Personalized Consumer Support	interactive personalization	Wang & Benbasat 2014; Johar et al. 2014
	webpage transparency	Lee & Benbasat 2011; Xu et al. 2014
	trust	Komiak & Benbasat 2006; Parboteeah et al. 2009

2.2 Web Personalization Research

Web personalization is a web-based information system that provides webpages adapted to the needs of individual users. Web personalization research provides three general web personalization approaches: record-based, interactive, or the combination of the two. In section 1.2, we briefly discussed record-based approach and interactive approach. Record-based approach utilizes the records of users' past web-search activities. Interactive approach mainly relies on runtime user-system interaction. Along with studying the three approaches, web personalization research has proposed a large number of techniques for processing user inputs and searching relevant webpages. In this section, we review major techniques of web personalization research. Section 2.2.1 reviews the major techniques for input processing. Section 2.2.2 reviews the major techniques for webpage searching.

2.2.1 Input Processing

Input processing extends user inputs in order to increase the accuracy in selecting relevant webpages (Furnas et al., 1987; Kumar et al., 2014; Manning et al., 2008; Song et al., 2014). In the literature, there are three types of input processing techniques, which are model-based, lexicon-based, and feedback-based. Most research proposes model-based technique while the smallest number of the works is feedback-based.

Model-based input processing utilizes user models derived from users' previous web-search activities. The major work includes Billerbeck et al. (2003), Cui et al. (2003), Gao et al. (2007), Yin et al. (2009), and Zhou et al. (2012). Recent improvement appears in Kumar et al.

(2014) and Kacem et al. (2014). The former provided Singular Value Decomposition (SVD) methods for social bookmarking services. The methods could produce a Clustered User Interest Profile (CUIP) model for each user. The CUIP was then used to expend user inputs to improve web personalization. Experimental evaluation supported that the proposed methods outperformed previous proposals. Kacem et al. (2014) proposed support vector machine (SVM) methods. They proposed an algorithm to incorporate context information in user inputs. Tegegne & Weide (2014) extended user queries by combining user preferences and existing knowledge. They proposed CP-nets that could capture user needs. Hahm et al. (2014) proposed using ontology to index engineering webpages and user models.

Early research in model-based input processing utilized search logs to produce user models. The methods matched search logs with user inputs (e.g., Huang et al., 2003, Yin et al., 2009). Most early research also used other information (e.g., search time) in search logs to produce user models (e.g. Beeferman & Berger, 2000; Billerbeck et al., 2003; Cui et al., 2003; Riezler et al., 2007). Some works built several user models for each user (Adomavicius et al., 2005; Koren, 2010). The other researches built a single model for each user (Koren et al 2009; Xiong et al. 2010).

Additionally, machine-learning or text-mining models are major players in input processing. The models incorporate previously visited webpages and user inputs. The models may evaluate the relevance of the current webpages to extend user inputs (Savoy, 2003; Shiri & Revie, 2003). The models provide a rank for the relevance of each webpage and select the most relevant webpages (Chang et al., 2006). Alternatively, some researchers manually constructed

input extensions that are stored in a repository. Relevant extensions will expand user inputs during users' web searches (e.g. Kraaij et al., 2003; Hu et al., 2006; Natsev et al., 2007; Park & Ramamohanarao, 2007; Turney et al., 2003). In a very different approach, Wu & Lin (2012) proposed the WNavis model by using link mining techniques. Liu et al. (2014) extracted the relevant terms from search results. The proposed model utilized the extracted entities to extend user inputs. Liu et al. also used user models. The experimental evaluation supported superior performance of the proposed solution comparing to the benchmark models.

Lexicon-based input processing utilizes synonyms and acronyms from external lexicon resources (e.g. thesaurus.com, WordNet, and Wikipedia) to expand user inputs. Domain knowledge may be extracted from external lexicon resources to expand user inputs (Gong et al., 2006; Navigli & Velardi, 2005; Nguyen et al., 2008; Tan & Peng, 2008). Often, heuristic rules are proposed to determine the similarity between user input and lexical term (Liu et al., 2004; Song et al., 2007). Besides lexicon assistance, syntactic analysis is also used (Sun et al., 2006; Liu et al., 2008; Song et al., 2006). In addition to online dictionaries, recent researches have utilized Wikipedia or other user-contributed knowledge base. Such resources are combined with anchor webpages constructed as standard webpages (Arguello et al., 2008; Kraft & Zien, 2004; Xu et al., 2009). Riezler et al. (2007) utilized Wikipedia FAQs. More recently, researches have been focused on user-system interaction. Evrim & McLeod (2014) utilized WordNet and domain ontologies to identify the changes of user inputs. The proposed Context-Based Information Analysis (CONIA) expanded user inputs with the terms from WordNet and domain ontologies.

Segura et al. (2014) utilized Gene ontology to expand user inputs. The researchers presented that the proposed method improved search accuracy significantly.

Feedback-based input processing provides feedback to the user-system interaction after the user submits the input. The feedback may ask questions or give suggestions to the user who can improve the initial input. In general, feedback-based input processing utilizes the techniques for other two types of input processing discussed above. For example, Rattenbury & Naaman (2009) contributed an algorithm to incorporate interactive activities during a web-search session. Rendle (2012) proposed context-aware input processing with factorization machines. The researcher demonstrated improved effectiveness and efficiency of the proposed solution by comparing to prior well-known method. Liu et al. (2014) proposed a web search system called EntEXPO. It identified key entities in user inputs and retrieved similar terms from knowledge bases. The retrieved terms expanded user inputs. The proposed method utilized the techniques from lexicon-based and model-based input processing.

2.2.2 Webpage Searching

Web-search methods have been proposed by a large body of researches. Most of the proposed approaches are ranking that is a process to rank the webpages by relevance. Lately, none-ranking approaches become popular. In this section, we review popular techniques of ranking and non-ranking approaches.

Ranking approach is most used in research and industry. A well-cited research, Page et al. (1999), described the PageRank used by Google Search. The algorithm exploited the link

structure of the Web and ranked every webpage on the Web. When a user submitted the input, the system evaluated the similarity between the input and each webpage. Then, the system ranked the webpages by relevance. Using a different ranking approach, Smyth & Balfe (2006) constructed usage matrices to record user inputs and user clicks on webpages. They proposed the I-SPY system to assess the similarity between two inputs, and then, rank the relevance of webpages. The system performed relevance assessment twice. The search results come from the second relevance assessment.

Other works provided more advanced ranking approaches. For example, Pitkow et al. (2002) adapted user inputs according to the user models. Then the system selected webpages utilizing the adapted inputs. In the same way, Liu et al. (2004) utilized Google Directory to adapt user inputs. The system then worked with Google to rank relevant webpages. Moreover, the user could improve their inputs during the web searches. The system could update the webpages on the basis of the changes in user inputs.

In a unique way, Bo et al. (2012) proposed ranking models and regularization algorithm that incorporated advanced predictive capability. The algorithm derived the prediction from a ranking model. The method imposed margin and slack constraints. In addition, the researchers proposed the adaption probability. Pana et al. (2013) proposed a graph-based regularization algorithm. The researcher evaluated the algorithm to show its capability to better meet users' needs comparing prior algorithm. Extending traditional ranking approach, Micarelli & Sciarrone (2004) proposed a filtering approach called WIFS. It filtered out the irrelevant webpages and presented the user with only those that met the similarity threshold.

Non-ranking approach focuses on webpage relevance evaluation and content recommendation, or both. Webpage relevance evaluation is typically to cluster webpages into small groups that the user can explore one group at a time. Content recommendation is generally to identify webpages most similar to user input. The user receives the selected webpages directly.

Most research has proposed algorithms for word selection from the webpages (Manning et al., 2008). The algorithms produce vector matrix for each webpage. Cosine similarity is often used to measure similarity between two webpages. Two popular researches, Zamir et al. (1999) and Osiński et al. (2004), contributed two well-cited methods. Zamir et al. (1999) proposed the Grouper interface to cluster search results. The Grouper utilized user search behaviors and incorporated the context of the web personalization. Osiński et al. (2004) proposed an algorithm called Lingo. It not only clustered retrieved webpages but also provided a description for each cluster. The description provided additional information to the users.

Alternatively, many researches are based on Support Vector Machine (SVM) techniques (e.g. Mecca et al., 2006; Ramage et al., 2009; Osiński & Weiss, 2005). The SVM is probabilistic algorithm to evaluate the relevance of webpages. The techniques utilize the keywords for probabilistic test with heuristic iterations. Osiński & Weiss (2005) proposed a text-segmentation heuristic to extract relevant webpages while eliminating irrelevant webpages. The experimental evaluation supported the effectiveness of the proposed solution. Mecca et al. (2006) offered an incremental method on the basis of their SVM models. The method used keywords to evaluate webpage relevance through multiple incremental steps. The researchers constructed the webpages for experimental evaluation of the proposed method. The experiments

concluded that the proposed method could achieve high accuracy. But, the proposed method still performed poorly when using other webpages. Ramage et al. (2009) proposed an algorithm that derived two SVM models, which worked together to assess webpage relevance. The researchers argued that the two models could reduce search errors by enforced evaluation power. The experimental resulted in improved accuracy comparing to other methods (Hofmann, 1999; Ramage et al., 2009).

Unlike other works that were based on conventional nature language processing, Cobos et al. (2014) proposed the method WDC-CSK to improve the descriptions of webpage clusters. They combined the cuckoo search meta-heuristics and k-means. The proposed method utilized cluster split and merge techniques to improve accuracy. Also, the research applied Balanced Bayesian Information Criterion as fitness function to improve webpage clustering. The experiments compared WDC-CSK with several well-accepted webpage clustering algorithms. In all the comparisons, the WDC-CSK performed better.

Webpage recommendation recommends webpages to users. Porcel et al. (2012) studied research knowledge transformation by using the Web. They proposed a hybrid fuzzy linguistic recommendation system that combined webpage clustering and collaborative clustering. Webpage clustering clustered the recommendations according to user inputs. Collaborative clustering clustered research resources for user models. The system switched between webpage clustering and collaborative clustering to improve clustering results. Also, the system could recommend specialized and complementary research resources for interdisciplinary research collaborations. Cobos et al. (2013) investigated web personalization for pedagogical lecture

development. They proposed the Recommendation System of Pedagogical Patterns (RSPP), including the ontology of pedagogical patterns and a hybrid recommendation method. The system used LSI to cluster lecture information on the Web. Similarly, Kardan & Ebrahimi (2013) also proposed a hybrid recommendation system for online discussion groups. They attempted to identify user postings. The proposed method used synonyms from WordNet and could provide more accurate recommendations than other hybrid systems.

There are other approaches for webpage searching. Steichen et al. (2009) pooled user responses during eLearning interactions. The pooled results might contain text, image, or video. In eLearning environment, the search spaces were often fixed specific webpages. Thus, pooling was very suitable without imposing unnecessary requirement to the user. The system displayed the recommended webpages in a user-friendly way. Steichen et al. (2011) evaluated the approach proposed by Steichen et al. (2009) in a field study. They implemented a prototype in a system of customer technical support. The prototype collected technical support webpages from web sources (online product's manuals, social networking, online discussion forums, etc.). The personalization method evaluated the user's knowledge and experience. The field study indicated the pooling technique proposed by Steichen et al. (2009) could improve system performance.

As presented at the beginning of this section, input processing and webpage search each has drawn a large body of researches. The review given above is less than exhaustive rather than providing an overview of prior research. For helping the readers navigate prior research, we present a brief summarization of input processing research and webpage search research in Table

2.

Table 2. Literature Review Summary (II)

Web Personalization Research		
Topics	Approaches	Example Researches
Input Processing	model-based	Gao et al. 2007, Kacem et al. 2014; Kumar et al. 2014; Yin et al. 2009, and Zhou et al. 2012
	lexicon-based	Evrin & McLeod 2014; Tan & Peng 2008; Xu et al. 2009
	feedback-based	Liu et al. 2014; Rattenbury & Naaman 2009; Rendle
Webpage Searching	ranking approach	Bo et al. 2012; Page et al. 1999; Pana et al. 2013;
	webpage relevancy	Cobos et al. 2014; Osiński et al. 2004; Zamir et al. 1999
	content recommendation	Kardan & Ebrahimi 2013; Porcel et al. 2012; Steichen et al. 2009

2.3 Review Helpfulness Research

Review helpfulness refers to the utility or usefulness of a review to help consumers make better purchase decisions. Review helpfulness is a synonym of review utility and review usefulness. Review helpfulness research provides novel solutions for utility-sensitive review analysis (USRA). Its objectives are to alleviate information overload to consumers and to solve the inherent problem of the helpful ranking used by review websites (Chen & Tseng, 2011; Ghose & Ipeirotis, 2007/2010; Liu et al., 2007; Ngo-Ye & Sinha, 2012).

In the literature, most review helpfulness research is aimed at finding effective approaches to predict review utility (Cao et al., 2011; Chen & Tseng, 2011; Ngo-Ye & Sinha, 2014). Prior research has contributed a large body of predictive models for automatic USRA. Majority research applies supervised machine learning techniques. Support Vector Machines

(SVMs) are the common method (Danescu-Niculescu-Mizil et al., 2009; Chen & Tseng, 2011; Kim et al., 2006; Ngo-Ye & Sinha, 2012; Zhang & Varadarajan, 2006). Alternative methods include decision tree (Liu et al., 2007; O'Mahony & Smyth, 2009), and clustering (Korfiatis et al., 2012; Mudambi & Schuff, 2010). Recently, we have seen models proposed for personalized utility-sensitive review analysis (PUSRA) (e.g. Moghaddam et al., 2012).

Moreover, natural language processing (NLP) and text mining play a key role in review helpfulness research (RHR). Lexical or syntactic NLP methods are most used and very effective (Chen & Tseng, 2011; Ngo-Ye & Sinha, 2014). Additionally, a few researches utilized semantic methods that may include ontology. For example, Cao et al. (2011) examined the semantic characteristics of online reviews.

Given the discussion above, we will review the existing researches in four subsections. Section 2.3.1 reviews the researches using supervised machine learning and the helpful ranks on review websites. Section 2.3.2 reviews the researches using supervised machine learning techniques but not using the helpful ranks. Section 2.3.3 reviews the researches using unsupervised machine learning techniques. Section 2.3.4 reviews the researches on PUSRA and presents the limitation of prior PUSRA approach.

2.3.1 Supervised Techniques with Helpful Ranks

So far, Support Vector Machines (SVMs) have been the most popular method employed by review helpfulness research (Ghose & Ipeirotis, 2007/2010, Kim et al., 2006; Zhang & Varadarajan, 2006). A large group of the researches study online reviews on electronic products.

Kim et al. (2006) proposed the SVM models to examine online reviews in multiple ways, e.g. structural, lexical, syntactic, and semantic. They found that the most effective NLP techniques were length-based and lexicon-based. The review helpful ranks closely related to the parse of speech and the characteristics of the products and the reviews. The most significant factors were review length, unigram characteristics, and star ranks. Particularly, the unigrams impacted the review utility greater than did the bigrams.

Like Kim et al. (2006), Zhang & Varadarajan (2006) used SVM techniques. But, Zhang & Varadarajan examined the online reviews using an extended method. Their model determined helpful reviews according to their similarity to the editorial reviews. Zhang & Varadarajan (2006) studied linguistic style but not the helpful ranks on the review websites. They tested the model by using the reviews at Amazon.com. Unlike Kim et al. (2006), Zhang & Varadarajan (2006) found parse of speech did not impact review helpfulness. Like Kim et al. (2006), Zhang & Varadarajan (2006) concluded that shallow syntax was highly influential on review utility. The most important factor was linguistic style cue. But, editorial reviews did not improve the accuracy of review helpfulness analysis. That finding is at least partially consistent with the finding by David & Pinch (2006). They concluded that editorial reviews were intended for sales promotion. Thus, they had less influence to consumers than online reviews.

Ghose & Ipeirotis (2007) studied the relationship between review subjectivity and review utility. The proposed model computed sentence-level probability, average subjectivity, and subjectivity deviation. The research found that subjectivity deviation and review readability significantly impacted review helpfulness. Ghose & Ipeirotis (2007) extended the finding of Kim

et al. (2006) and Zhang & Varadarajan (2006). Later, Ghose & Ipeirotis (2010) extended the work of Ghose & Ipeirotis (2007). The former proposed the SVM models with additional evaluations of reviews and reviewers. They extracted the reviewer profiles from the review websites. The research studied the readability of online reviews by combining reviewer profiles.

The recent improvement in review helpfulness research is to incorporate the social contexts of the reviewers. Lu et al. (2010) found that the information of the reviewer's social context can enhance the predictive accuracy of review utility. They argued that the social context indicated the quality of a reviewer, which in turn influenced the quality of the reviews written by the reviewer. They postulated four consistencies: author, trust, co-citation, and link. The author consistency means that the reviews from the same author are of similar quality. The trust consistency means a social link from reviewer A to reviewer B implies trust relationship between them. Reviewer A trusts reviewer B only if the quality of reviewer B is at least as high as that of reviewer A. The co-citation consistency means that people are in common regarding the way to trust the others. Therefore, if review C trusts reviewers A and B, three of them have similar quality. The link consistency means if two reviewers are connected in a social network, then their reviews are the same quality. Lu et al. (2010) used the four consistencies to test the proposed model. They evaluated it by using reviewer profiles from the review website, Ciao (www.ciao.co.uk). It is a community-based review website where the reviewers also rank the reviews written by the other reviews. At ciao.co.uk, one can form one's own Circle of Trust and add trusted members to it. Clearly, the proposed approach for review helpfulness analysis is suitable only for the review websites that support trust network.

Ngo-Ye & Sinha (2012) compared their proposed model with other models for USRA. The researchers used the raw number of helpful votes downloaded for the review website. For example, if the review website provides the summary “x of y people said the following review helpful,” then the “x” is the helpfulness rank of a review. The larger is the number x, the more helpful is the review. They proposed the feature selection method called Regressional ReliefF (RReliefF). The researchers compared their model to the baseline BOW model. The result of the comparison was that their model increased the accuracy of USRA substantially. Moreover, comparing to the other methods, the RReliefF is simpler and more efficient. The RReliefF could thus reduce workload while incorporating the contextual information in the k nearest neighbors.

Following from Ngo-Ye & Sinha (2012), Ngo-Ye & Sinha (2014) found significant effect of review webpages and reviewers characteristics on review helpfulness. Ngo-Ye and Sinha used a baseline model to examine the effective of the proposed Vector Space Model (VSM). The researchers compared the performance of their model with the baseline model. The evaluation indicated that the proposed model could substantially improve the accuracy of utility-sensitive review analysis (USRA) comparing to the baseline model.

As mentioned before, some researches have studies the semantics of online reviews. Chen & Tseng (2011) and Cao et al. (2011) both studies review semantics to predict review utility. The difference between the two works is that Chen & Tseng (2011) did not use the helpful ranks on review website, but Cao et al. (2011) utilized them. We will review Chen & Tseng (2011) in Section 2.3.2. Cao et al (2011) examined online reviews in three aspects: basic, stylistic, and semantic. They studied a large crop of online reviews regardless their helpful ranks

on a review website. They compared different combinations of review helpful ranks. They found that the semantics of online reviews were more influential than the syntax and style.

Besides the works discussed in previous paragraph, other researches have contributed unique findings. Hoang et al. (2008) and Otterbacher (2009) found that writer authority (e.g. the reputation of the reviewer) impacted review helpfulness. Otterbacher (2009) used the data quality framework proposed by Wang & Strong (1996). Otterbacher found that the reviewer's reputation was a primary factor of review quality. Similarly, O'Mahony & Smyth (2010) discovered that the rank of a reviewer greatly affected review helpfulness. In general, empirical studies support the impact of reviewer's reputation on review utility (Cheung & Lee, 2012). Ku et al. (2012) explained that high reputation of a reviewer suggested the trustworthy quality of the reviews written by that reviewer.

2.3.2 Supervised Techniques without Helpful Ranks

The helpful ranks on review websites may not be used to build predictive models for USRA. O'Mahony & Smyth (2009) examined reviewer's reputation and social context as well as review webpage and sentiment. The reputation of a reviewer was indicated by the rank received by the reviewer. The review features included the length of a review and the ratio of uppercase to lowercase characters. The social features mainly involved the number of reviews written by a reviewer and the mean of review quantities.

Liu et al. (2007) provided a model that subdivided the reviews into best, good, fair, and bad in terms of review utility. The model evaluated richness and persuasiveness of online review.

The researchers did not use the helpful ranks on the review websites because of the problem in the helpful ranks. The researchers found that many reviews did not receive enough helpful votes to be ranked as highly helpful. The researchers manually selected the set of helpful reviews as gold review set that was the set of reviews regarded as helpful by the researchers. Against the gold review set, the researchers tested the predictive accuracy of the proposed model evaluating three aspects: informativeness, subjectiveness, and readability. Each of the three aspects consists of sub-divided factors (e.g. product-aspect mention, pattern of review titles, sentence sentiment polarity, and paragraph structure). The research found that title-appearance does not affect the performance of review utility. In the same fashion, Liu et al. (2012) proposed the model of USRA for product designers. The researchers used manually selected helpful reviews from amazon.com as gold review set. They examined review texts and proposed the predictive model that could help product designers evaluate review helpfulness.

Chen & Tseng (2011) applied the information quality framework proposed by Wang & Strong (1996). Chen and Tseng used expert-constructed reviews as the gold review set. Against it, they tested their proposed model. The researchers proposed five review-quality levels: high, medium, low, duplicate, and spam. Multiclass support vector technique was used in the model. They found that the most helpful reviews were those giving rich and in-depth opinions. In addition, Chen & Tseng (2011) proposed the method to mining the opinions of the reviews. The proposed method could not only predict the helpfulness of a review but also incorporate personalized review helpfulness analysis.

2.3.3 Unsupervised Techniques

In review helpfulness research, the unsupervised techniques are mainly clustering method. Tsur & Rappoport (2009) studied the helpfulness of book reviews. They identified a set of helpful reviews and compared the unhelpful reviews against the helpful reviews. Accordingly, they used the selected helpful reviews as the gold review set and determined the utility of each review according to the similarity between unhelpful review and helpful review.

Alternatively, a few researches used gold review set in a different manner. Lappas & Gunopulos (2010) selected a small set of helpful reviews that gave rich opinions. They summarized the helpful reviews and used their summary to evaluate the utility of an online review. In a similar approach, Tsaparas et al. (2011) proposed an algorithm that could select a comprehensive set of helpful reviews that were summarized to produce an opinion-rich review. The researchers argued that their approach could maximize the coverage of helpful reviews.

In a unique approach, Korfiatis et al. (2012) studied the interrelationships among review helpfulness, helpful ranks, and readability of online review. They proposed a model that explained the effect of conformity, understandability, and expressiveness. The research found that the readability of online review had the most significant impact on the helpfulness ranks. Writing style and text length also affected review utility. Drawing the similar finding as that drawn by Mudambi & Schuff (2010), Korfiatis et al. (2012) concluded that product rating affected the helpful rank. Positive reviews tended to be long text while negative reviews tend to be short text.

2.3.4 Personalized Review Helpfulness

In Section 2.3.2, we mentioned that Chen & Tseng (2011) proposed a method that could perform personalized utility-sensitive review analysis (PUSRA). However, their focus was on particular groups of hotel consumers. The work did not extend to PUSRA for individual consumers. In the literature, Moghaddam et al. (2012) is probably the only research that proposes PUSRA models for individual consumers. In Section 1.2, we presented that vast majority review helpfulness research assumes common utility. But, Moghaddam et al. (2012) assumed personal utility and challenged common utility assumption. They argued that the utility of a review might vary from one consumer to another. Therefore, utility-sensitive review analysis (USRA) should be similar to personalized recommendation.

Moghaddam et al. (2012) utilized the helpful ranks on review websites to examine reviews, reviewers, consumers, and products. The proposed models involved the characteristics of reviews, consumers, and products to determine the utility of the review. In order to provide personalized recommendation, they proposed a series of probabilistic models. Their experiments demonstrated that the helpfulness of a review indeed differed from one consumer to another.

However, there are two major weaknesses related to the models proposed by Moghaddam et al. (2012). First, many researchers have observed the inherent problem of the helpful ranks on review websites. Involving them in the model may reduce the validation of the research solution. Second, the paper did not offer a web personalization artifact required from PUSRA (see Section 1.2).

In summary, prior review helpfulness research has made enormous contribute to USRA.

The main focuses of prior USRA model are summarized in Table 3.

Table 3. Literature Review Summary (III)

Review Helpfulness Research (RHR)		
Approaches	Main Focuses	Example Researches
Supervised Techniques with	review length, syntactic feature, & semantic feature	Cao et al. 2011; Ghose & Ipeirotis 2007/2010
	review linguistic style	David & Pinch 2006; Zhang & Varadarajan 2006
	review subjectivity	Ghose & Ipeirotis 2007/2010
	reviewer trust, link, & reputation	Lu et al. 2010; Ngo-Ye & Sinha 2012/2014
	review quality	Otterbacher 2009
Supervised Techniques without	reviewer characteristics, e.g. social context	O'Mahony & Smyth 2009
	opinion richness	Liu et al. 2007/2012
	review quality	Chen & Tseng 2011
Unsupervised Techniques	opinion richness	Lappas & Gunopulos 2010; Tsaparas et al. 2011
	product rating	Korfiatis et al. 2012; Mudambi & Schuff 2010
	review readability	Korfiatis et al. 2012
	gold review set	Lappas & Gunopulos 2010; Tsur & Rappoport 2009
Personalized Review Helpfulness	utility-sensitive review analysis for similar consumers	Chen & Tseng 2011
	utility-sensitive review analysis for individuals	Moghaddam et al. 2012

CHAPTER 3 RESEATCH METHODOLOGY

Information Systems research (ISR) encompasses dual paradigms, behavioral science (BS) and design science (DS). BS is aimed at explaining and predicting phenomena related to information systems (IS) artifacts. DS is aimed at creating novel and useful IS artifacts to solve unsolved problems facing people or organizations (Hevner et al. 2004; March & Smith 1995; March & Storey 2008). The term artifact refers to man-made object for achieving human purpose; that meaning was given by Herbert A. Simon in his book, ‘Sciences of the Artificial’ (Simon 1996). He distinguished natural objects from artificial objects. The former refers to natural objects while the latter refers to man-made objects. In IS field, information systems and their components are artifacts that are designed for enhancing effectiveness and efficiency of human activities (Baskerville et al., 2011; Hevner et al., 2004; March & Smith 1995). According to Hevner et al. (2004), there are four types of IS artifacts: “constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), and instantiations (implemented and prototype systems) (Hevner et al., 2004, pp. 3).”

Simon (1996) legitimized design science (DS) being a scientific paradigm. It is a problem solving and a legitimate research paradigm in ISR (Hevner et al., 2004). Additionally, ISR has another scientific paradigm behavioral science (BS). Its root is in natural science. BS seeks to explain the phenomena surrounding managing and using IS artifacts. In summary, ISR possesses the dual paradigms, DS and BS. They both are fundamentals of ISR. They are also complementary and interdependent (Hevner et al., 2004; March & Storey, 2008).

Our current work is a DS. Our research methodology complies with the DS framework and guidelines proposed by Hevner et al. (2004). Their contributions have become the most applied DS methodology that guides DS researches to achieve both relevance and rigor (Goes, 2014). Relevance requires DS artifacts be useful and effective in achieving the specific human or organizational goals. Rigor requires DS research appropriately utilize prior knowledge in the body of knowledge base. Moreover, Hevner et al. (2004) highlighted the requirement for DS research to contribute to the body of knowledge base, i.e. theoretical contribution. The requirement is consistently held by prominent works in DS methodology (e.g. Baskerville et al., 2011; Goes, 2014; Kuechler & Vaishnavi, 2008; March & Smith, 1995; Vaishnavi & Kuechler, 2011). In this research, we hold theoretical contribution as a major research objective. In order to achieve it, we appropriately apply Hevner et al. (2004)'s guidelines. By doing so, we present our innovative and effective design-science artifacts for DS theoretical development.

This chapter illustrates the research methodology applied in this paper in four sections. Section 3.1 compares DS and BS as well as presents their interrelation. In section 3.2, we review the major literature of DS methodology in Information Systems (IS) field. In Section 3.3, we discuss our research methodology and our application of Hevner et al. (2004)'s design science guidelines that guide this research to achieve research rigor and relevance.

3.1 Behavioral Science and Design Science

Behavioral science (BS) research in Information Systems (IS) is aimed at explaining and predicting phenomena. BS develops theories that provide understanding of organizational

phenomena surrounding IS artifacts (Hevner et al., 2004; March & Smith, 1995). BS theories inform IS researchers and practitioners with the understanding of managing and using IS. The understanding can inform better design of DS artifacts to achieve their intended purposes (Hevner et al., 2004). In contrast, design science (DS) is intended to extend human capability and to help organization solve problems. DS innovates DS artifacts for effective and efficient analysis, design, implementation, management, and use of information systems. DS artifacts are directly relevant to the practices of IS professionals. The relevance of DS theories is the direct applications of the novel and useful DS artifacts in business practices (Hevner et al., 2004).

There are major differences between BS and DS. First, they are rooted in different scientific paradigms. BS has its root in natural science while DS has its root in engineering and artificial science (Hevner et al., 2004; Baskerville et al., 2011). Second, BS and DS have distinct goals. BS focuses on knowledge discovering by building and testing explanatory or predictive theories whereas DS concentrates on problem solving by innovation and creativity from which DS artifacts present design-science (DS) theoretical contributions (Hevner et al., 2004; Simon, 1996; Vaishnavi & Kuechler, 2011). Third, BS and DS relate to professional practices in different ways. BS is indirectly related to IS practices. The practical relevance of BS theories relies on effective DS artifacts (Markus et al. 2002, pp. 173). Although BS theories can inform IS and business professionals, the theories are most often unable to do so. That is because BS theories frequently lag behind new and effective design artifacts (Hevner et al., 2004; Markus et al., 2002, pp. 180). In contrast, the relevance of DS is direct provided by DS artifacts. They often

play a key role in industry innovation. They have typically passed through extensive utility evaluation to ensure practical relevance.

In spite of the difference shown above, BS and DS are interrelated. They are complementary and inseparable in Information Systems research (ISR). Its relevance requires seamless integration of BS and DS (Hevner et al., 2004; March & Smith, 1995; Orlikowski & Iacono, 2001). And, IS researchers need to engage “the complementary research cycle between design science and behavioral science to address fundamental problems faced in the productive application of information technology (Hevner et al., 2004 pp. 2, 3).” March & Smith (1995) illustrated that the complementary ISR cycle composed (1) the proactive DS to create DS artifacts to solve organizational problems, and 2) the reactive BS to explain or predict the phenomena around DS artifacts.

3.2 Influential Literature of Design Science Methodology

In information systems research (ISR), design science (DS) is a relatively new paradigm comparing to behavioral science (BS). Hevner et al. (2004) was the first research that provided methodological framework and guidelines for DS (Kuechler & Vaishnavi, 2008). Before Hevner et al. (2004), DS methodologies had been discussed by IS researches including March & Smith (1995) and Swaab et al. (2002). Since Hevner et al. (2004), prominent DS methodology contributions have made various improvements to the work of Hevner et al. (2004). This section provides an overview of the influential DS methodology contributions.

Prior DS methodology research has proposed different frameworks and principles for good DS researches. Hevner et al. (2004) addressed those issues comprehensively by proposing a framework and seven pragmatic guidelines. Their contributions cover how to execute and evaluate DS researches. The comprehensive framework and seven guidelines have become methodological basis “to assist researchers, reviewers, editors, and readers to understand the requirements for effective design-science research (Hevner et al., 2004, pp.82).” The proposed framework clearly defines the boundaries of DS in ISR. The proposed guidelines prescribe the practices of sound DS researches as well as DS artifacts to make DS theoretical contributions. Additionally, some researches regarded DS as pragmatic camp. Its counter party is the theorizing camp that weighs theorization more than design practice (Markus et al., 2002; Walls et al., 2004). However, most IS methodology researchers agree with Hevner et al. (2004) in that DS artifacts require creativity not clearly understood by the researchers. Some researchers have reconciled the two camps (Kuechler & Vaishnavi, 2012). Nevertheless, we agree with the viewpoint of Hevner et al. (2004) and the literature around their work.

The most cited seminal work is March & Smith (1995). They proposed a framework and research agenda for DS researches. Unlike Hevner et al. (2004), March & Smith (1995) sought to reconcile the conflict viewpoints over whether DS is legitimate scientific research in ISR. Through extensive demonstration, March and Smith established the scientific status of DS in IS field. They also proposed a research agenda that demonstrated the interrelation between DS and BS. However, March & Smith (1995) did not provide comprehensive guidelines for good DS research. Hevner et al. (2004) extended March & Smith (1995)’s research agenda with two

components, environment and knowledge base. Hevner (2007) proposed the framework of DS three-cycle to emphasize the importance of environment and knowledge base in DS research. Moreover, Hevner (2007)'s three-cycles framework illustrates that the framework and guidelines in Hevner et al (2004) is the methodological base for DS research to contribute DS theories. Alternatively, Swaab et al. (2002) provided a visualization method to present the creativity of DS researchers. Their contribution illustrates the indispensable role of researchers' creativity to successful DS research.

Recently, Vaishnavi & Kuechler (2011) extended Hevner et al. (2004) in several ways. Vaishnavi and Kuechler clarified the difference between DS research and design research. Design research is the study of design itself and designers. Design research is research into or about design. In contrast, DS research is the research with design, "using design as a research method or technique (Vaishnavi & Kuechler, 2011, pp. 3)." Also, Vaishnavi & Kuechler (2011) added a fifth type of IS artifacts, namely better theories. The fifth type gives some elaboration on presenting DS artifacts as DS theoretical contribution. The researchers argued that good DS researches could contribute to better theories through demonstrating rigorous design activities to create the artifacts. Further, good DS theories should illustrate the interrelationship between solution components. And, good DS theories should provide community-determined outputs (e.g., human-computer interface).

In addition, there are other DS methodology contributions that focus on balancing rigor and relevance in DS research. Prominent works include Baskerville et al. (2011). They stressed the term design as a verb in DS research. The researchers believed that DS research had

overstated research rigor while neglecting designers' creativity. The researchers called for a duality of design and science in DS research. Baskerville et al. (2011) argued that DS research should go back to the balance emphasized by Hevner et al. (2004). Baskerville et al. (2011) contributed three guidelines: (1) research domain as design domain; (2) understanding the domain as research objective; (3) theory contribution as primary research goal. In the same vein, Alter (2013) stressed the long-term research activities to produce work system method and work system theory. Ultimately, DS literature uniformly advocates both useful artifacts and DS theoretical contributions.

The previous discussion of prior DS methodology indicates that Hevner et al. (2004) provides the most used methodological base for DS research. The comprehensive framework and guidelines make them a practical method for DS research. Therefore, our current research applies Hevner et al. (2004)'s framework and guidelines as research methodology. We believe that appropriate applications can ensure our research to make significant DS theoretical contribution with both research relevance and research rigor. Hevner et al. (2004)'s guidelines establish the triple cycles among relevance, rigor, and theoretical contribution (Baskerville et al., 2011; Hevner, 2007). Using Figure 2 of Hevner et al. (2004), the following Figure 2 depicts their DS framework that shows the triple cycles:

“The contributions of behavioral science and design science in IS research are assessed as they are applied to the business need in an appropriate environment and as they add to the content of the knowledge base for further research and practice. A justified theory

that is not useful for the environment contributes as little to the IS literature as an artifact that solves a nonexistent problem (Hevner et al., 2004, pp. 81).”

Hevner et al. (2004) setup a solid basis from which many DS methodology researches sprout. Over the last ten years, Hevner et al. (2004)’s framework and guidelines have become a de fact to DS methodology emphasized by prominent DS methodology research (e.g., Baskerville et al., 2011; Kuechler & Vaishnavi, 2012). Our current work contributes to Hevner et al. (2004)’s applications by producing high-quality DS artifacts and DS theoretical contributions.

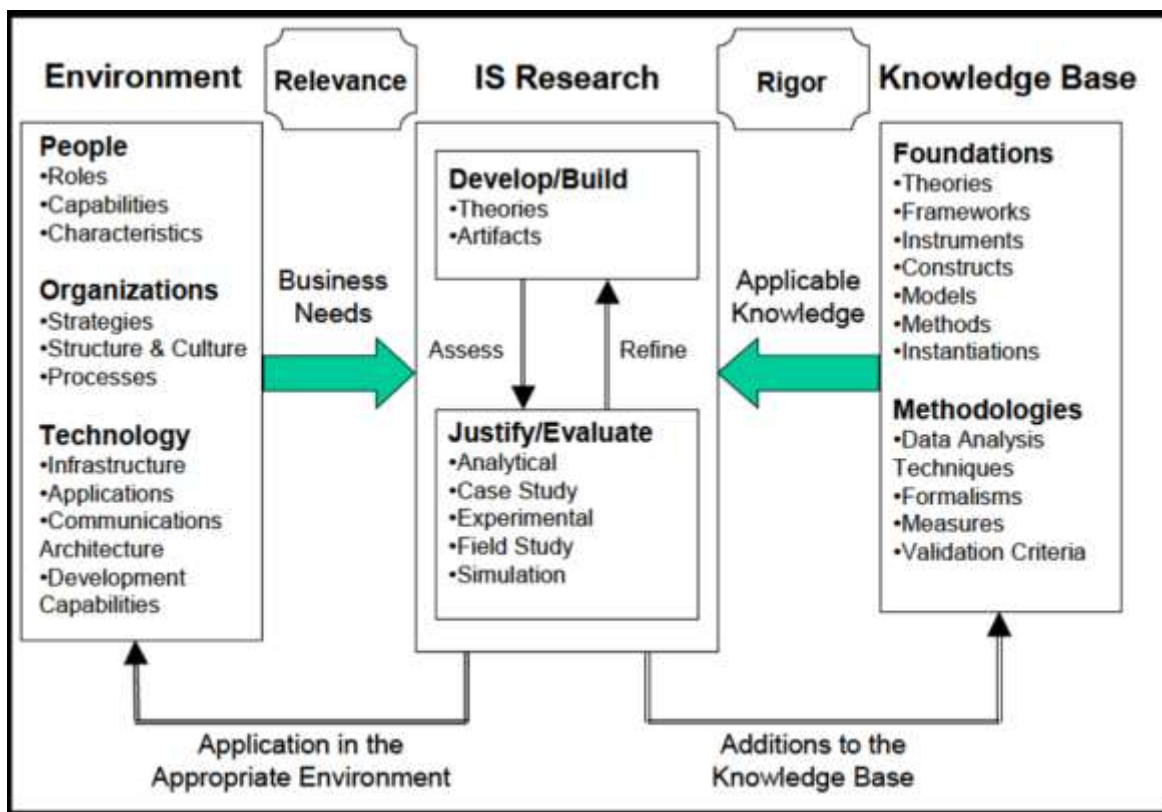


Figure 2. Design Science Framework by Hevner et al. (2004)

3.3 Our Research Methodology

Our research follows design science (DS) paradigm. Particularly, we apply Hevner et al. (2004)'s seven guidelines as methodological standard. This section outlines Hevner et al. (2004)'s seven guidelines and our application of them. Hevner et al. (2004) argued that a DS research should address each of the seven guidelines in some way in order to be complete. Therefore, our research applies each of them appropriately.

Guideline 1: Design as an Artifact

Hevner et al. (2004) emphasized the essential role of novel, purposeful, viable DS artifacts of four types: construct, model, method, or instantiation (Hevner et al., 2004). DS research must create DS artifacts since DS knowledge is acquired through creating and applying useful DS artifacts. A DS research needs to create artifacts that are novel and effective to solve the research problem. The descriptions of the design artifacts must enable effective implementations and applications in suitable application domains.

In our research, we create innovative artifacts of all four types via an iterative design process. The propose utility-sensitive review analysis (USRA) artifact is a web-based information system. It is an overarching design science artifact consisting of construct, model, method, and instantiation. Our proposed artifacts of four types are novel and useful to provide personalized utility-sensitive review analysis (PUSRA). The prototype has been evaluated and demonstrated the effectiveness of our solution. We present our proposed artifacts to broad

audience including IS researchers, IS professionals, and business professionals. Our proposed solution is applicability in diverse information system problems related to PUSRA.

Guideline 2: Problem Relevance

Hevner et al. (2004) stated that design science is a problem-solving paradigm. DS research is aimed at developing innovative artifacts to solve important yet unsolved business problems. The research problem of DS research defines the goals of the future information systems, which need to be reached from the current systems. Successful design artifacts help IS professionals meet those goals. Goal achievements will help solve business problems and reach business objectives (e.g. increasing revenue, increasing competitive edge, and reducing cost). So, DS research problem defined for attaining the goals must be relevant to IS professionals who attempt to solve their unsolved problems that impede the achievement of business objectives.

Our research investigates the unsolved, difficult problem of inadequate approach for utility-sensitive review analysis (USRA) concerned by IS professionals, specifically the designers of websites. The designers need to address the inherent problem of review helpful ranking. They need better DS artifacts to eliminate the problem of missing vote. Although a large body of researches has proposed useful models for USRA, prior research contribution fails to provide the needed DS artifacts, i.e., web personalization artifacts. No such DS theoretical guidance has been proposed. The existing models of USRA do not provide suitable solutions meeting practical need of PUSRA although the models have made substantial progress for USRA. Our research addresses the unmet need for effective PUSRA solutions. Therefore, our

research is highly relevant not only to IS researchers and IS professionals but also to business professionals and their organizations.

Guideline 3: Design Evaluation

Hevner et al. (2004) regarded solution evaluations as crucial components of design science (DS) research. The evaluations need to apply appropriate evaluation methods to demonstrate rigorously the utility and efficacy of the DS artifacts. DS researchers need to define appropriate measures for the evaluations in order to show the effectiveness of the DS solution. Hevner et al. (2004) suggested five sets of evaluation methods that we present in Figure 3 mirroring Table 2 of Hevner et al. (2004).

According Hevner et al. (2004), we apply three evaluation methods, namely analytical, descriptive, and experimental (see Figure 3). The combination of the three methods can comprehensively evaluate our DS artifacts in term of utility and efficacy. In the analytical method, we conduct detailed examination the structure of our design artifact in order to assess its quality. In descriptive evaluation, we rigorously apply relevant web personalization knowledge in the knowledge base. We provide illustrative use case (section 4.2) to present the internal consistency of our proposed solution. Moreover, we conduct two experiments to comprehensively evaluate the practical utility and efficacy of the proposed solution. Our evaluation contributes what we believe, the first efficiency evaluation of USRA DS artifacts to review helpfulness research. Thus, the evaluation results provide more reliable evidences for

practical usefulness and effectiveness of our proposed solution. We discuss the two experiments in detail in Chapter 8 USRAnalyzer Evaluation.

1. Observational	Case Study: Study artifact in depth in business environment
	Field Study: Monitor use of artifact in multiple projects
2. Analytical	Static Analysis: Examine structure of artifact for static qualities (e.g., complexity)
	Architecture Analysis: Study fit of artifact into technical IS architecture
	Optimization: Demonstrate inherent optimal properties of artifact or provide optimality bounds on artifact behavior
	Dynamic Analysis: Study artifact in use for dynamic qualities (e.g., performance)
3. Experimental	Controlled Experiment: Study artifact in controlled environment for qualities (e.g., usability)
	Simulation – Execute artifact with artificial data
4. Testing	Functional (Black Box) Testing: Execute artifact interfaces to discover failures and identify defects
	Structural (White Box) Testing: Perform coverage testing of some metric (e.g., execution paths) in the artifact implementation
5. Descriptive	Informed Argument: Use information from the knowledge base (e.g., relevant research) to build a convincing argument for the artifact's utility
	Scenarios: Construct detailed scenarios around the artifact to demonstrate its utility

Figure 3. Design Evaluation Methods Suggested by Hevner et al. (2004)

Intuitively, observational method evaluates DS artifacts in real-world environment, and thus, the method is important for assessing practical utility and efficacy. Also, testing method is probably the most effective method for performance evaluation. However, observational method is not suitable for our evaluation since the method does not support generalizable evaluation of personalized USRA within acceptable costs. Moreover, testing method is appropriate to evaluate DS artifacts whose practical usages are limited. But, USRA personalization has ever-extended usages that make testing method is inappropriate.

Guideline 4: Research Contributions

A DS research should contribute DS artifact, design knowledge, and/or evaluation knowledge; at least one type of DS contributions must be found and clearly presented in a given research (Hevner et al. 2004). Our research offers a novel personalized utility-sensitive review analysis (PUSRA) artifact as an overarching DS artifact that consists of constructs, models, methods, and instantiation. Our proposed artifact contributes to the knowledge base of review helpfulness research (RHR) with effective USRA approach. Additionally, our research contributes new knowledge to web personalization research (WPR). Its existing approach can be improved by our DS knowledge. Most importantly, we contribute the first comprehensive evaluation including effectiveness and efficiency to RHR and WPR.

Guideline 5: Research Rigor

A DS research needs demonstrate its rigor by suitably applying prior knowledge in the knowledge base. It includes the knowledge bodies of behavioral science and design science. Prior knowledge also includes appropriate evaluation methods (Hevner et al. 2004). For research rigor, we exploit the existing literature in web personalization. The literature has accumulated a large body of knowledge about the effective design approaches of web personalization systems. As discussed in Guideline 3, we will properly apply three types of the evaluation methods in the knowledge base. When exploiting evaluation methods, we carefully assess the appropriateness of prior knowledge for our current research.

Guideline 6: Design as a Search Process

Hevner et al. (2004) presented that DS research is a problem-solving process and relies on designing IS artifacts. The design process involves trials-and-errors. During our design innovation, we create design ideas and try them in our lab. We constantly identify deficiencies in the trial results, and then, improve the design components.

In such way, our current research has gone through numerous trials-and-errors. Due to the novel nature of our DS artifacts, the knowledge base discussed above is not adequate to a large extent. Thus, we heavily rely on our creativity to generate design ideas. Each of them is intensively tested in our lab. So far, many early ideas have proven unsuitable and have been rejected. After many trials-and-errors, we have fundamentally improved the original design. The solution components presented in this paper have proven the best testing results. After those arduous works, we have proposed and prototyped our DS artifacts that Chapter 4 through Chapter 6 will present.

Guideline 7: Communication of Research

DS researchers need to present their researches suitably for technical and business audiences. The researchers need to provide good presentations so that the proposed DS artifacts can be applied by IS researchers to extend research contributions and by IS professionals to solve organizational problems (Hevner et al. 2004). In this paper, we strive to communicate with broad audiences, specifically IS researchers, IS professionals, business professionals, and consumers.

CHAPTER 4 MODEL AND ARCHITECTURE

In section 3.4, we presented Hevner et al. (2004)'s design science (DS) framework and guidelines. The Guideline 6 suggests design as a search process. After numerous trials-and-errors, we have created an effective solution for personalized utility-sensitive review analysis (PUSRA). Our proposed solution is an interactive web personalization (WP) artifact. We refer to it as USRAnalyzer. In chapter 1 and chapter 2, we discussed WP artifacts, which is a web-based system designed to provide personalized webpages meeting the needs of individual consumers.

Moreover, WP artifact is two types: record-based or interactive. The former requires user models on the basis of users' past web visits. The latter does not requires user models but rather using users' runtime activities. We propose the USRAnalyzer as an interactive WP artifact for three reasons. The first reason is interactive WP artifact can better satisfy consumer needs than can record-based (see section 2.2). Another reason is that the users of the USRAnalyzer are most often first-time users. So, it is impossible to derive the user models for them. The third reason is the USRAnalyzer as an interactive WP artifact can better address the changes of consumer needs during their review searches. Therefore, the USRAnalyzer can be more effective to provide dynamic PUSRA.

In this chapter, we describe our proposed USRAnalyzer overview model (section 4.1) and architecture (section 4.3). Also, we offer a use case in section 4.2 to demonstrate how the USRAnalyzer can offer PUSRA.

4.1 USRAnalyzer Overview Model

Prior solution for PUSRA provides predictive model (see section 2.3.4) which is generally based on the users' past activities and personal characteristics. The model is unable to meet the changes of consumer needs during the process of making purchase decisions. In contrast, the USRAnalyzer is an interactive consumer decision support system that can interact with a consumer and meet the changes of consumer needs. The USRAnalyzer helps consumers make better purchase decisions with consumer-centric design. The functionality of the USRAnalyzer meets interactive web personalization (WP) requirements, delivering to consumers an advanced WP. In Figure 4, we illustrate the overview model of the USRAnalyzer. We discuss its major functions as following.

We regard a helpful review as an online review relevant to a specific need of a consumer during his/her purchase decision. During a session of consumer decision support, the consumer inputs his/her need into the USRAnalyzer (box 1 of Figure 4). The consumer input c_j is considered the best expression of the consumer's need (Davern & Kamis, 2010; Häubl & Murray, 2006; Lappas & Gunopulos, 2010; Xu et al., 2014). The USRAnalyzer receives c_j and performs input processing on it (box 2). We will describe the input processing in section 4.3. In general, the c_j is a natural language text in some natural language (English, Chinese, German, Japanese, etc.). The USRAnalyzer also expands c_j with synonyms and acronyms (see section 4.3).

After input processing, the USRAnalyzer conducts the initial relevance evaluation of the reviews in the review pool RP on the review website (box 3). The RP contains all the online reviews of the product/service interesting to the consumer. The USRAnalyzer can access the RP by connecting to its link provided by the website. The initial relevance evaluation is aimed at identifying the relevant review $r_j \in RP$ because in RP , there are typically many reviews that are irrelevant to the consumer input c_j . Intuitively, the USRAnalyzer needs to reject the irrelevant reviews because they are not helpful to the consumer.

Identifying relevant reviews at box 3 usually does not provide the final results. Most often, the reviews in RP have different relevance levels. A review in RP is a relevant review $r_j \in RP$ if it is similar to c_j . High similarity between them renders a high relevance level (l_j) of r_j . As discussed previously, some reviews in RP are irrelevant to c_j and should not be considered by the USRAnalyzer.

After rejecting the irrelevant reviews, the set of relevant reviews is the *relevant set* (RS_j), which is a sub-set of RP . The RS_j contains all relevant reviews. When the size of the RS_j is large, a consumer generally wants to read the highly-relevant reviews satisfying a *desirable relevance level* (d_j) determined by the consumer. Therefore, at box 4, the *USRAnalyzer* allows the consumer selecting the d_j from a list of relevance levels (e.g. very high, high, medium, and low). The consumer's selection is regarded as a very useful way to interact with the consumer for accuracy helpful review provision (Reisen & Hoffrage, 2010). Thus, the *USRAnalyzer* suggests the consumer selecting the d_j from the predefined relevance levels. We use the commonly-used four levels: *very high, high, medium, and low*) (refer to section 4.2).

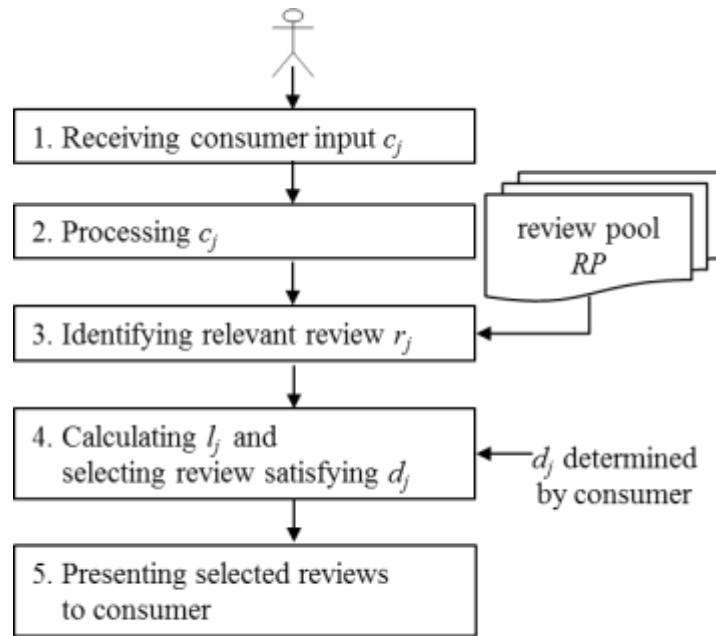


Figure 4. USRAnalyzer Model

After receiving the consumer’s selection, the USRAnalyzer evaluates the l_j of each relevant review in RS_j , and selects the reviews satisfying $l_j \geq d_j$. Further, the USRAnalyzer ranks the reviews by relevance level. Finally, the USRAnalyzer presents the consumer with the selected reviews sorted in descending relevance level (box 5).

After the final result, the consumer may change the input and start a new session of consumer decision support. In the new session, the interaction between the consumer and the USRAnalyzer is the same as that described above.

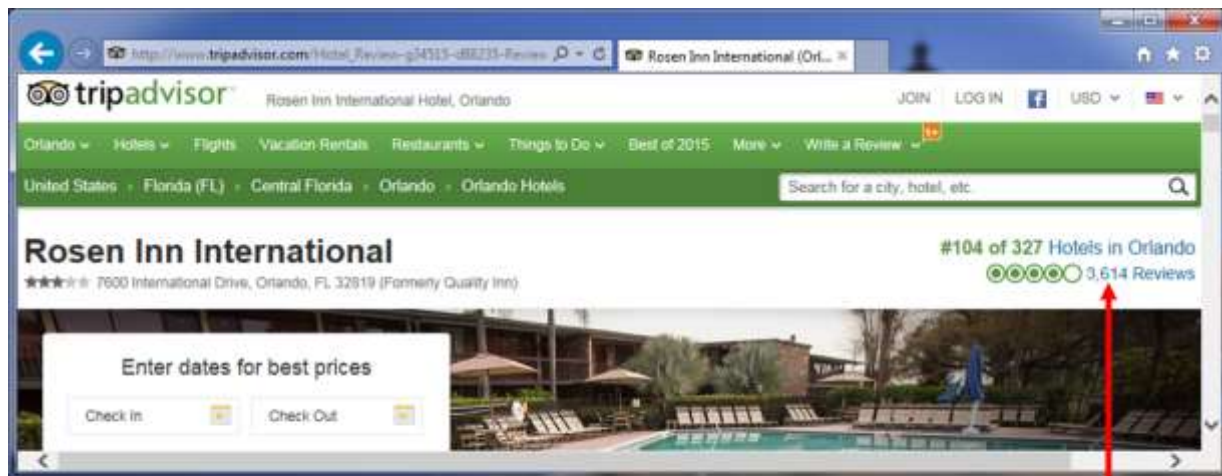
4.2 A Use Case of USRAnalyzer

This section describes a use case where a fictional consumer Kin interacts with the USRAnalyzer to select the relevant reviews in order to plan her trip to Orlando, Florida, USA.

The use case illustrates how the USRAnalyzer satisfies Kin's need with helpful reviews identified from a large review pool RP. We use real-world online reviews and actual screen shots produced by our USRAnalyzer prototype. The screen shots visualize the viability and efficacy of the USRAnalyzer.

The use case: A consumer Kin is making a hotel-reservation decision: which hotel in Orlando, Florida, USA she will make a reservation for her trip to Orlando. She is very interested in Rosen Inn International (RII) in Orlando and wants to see how the online reviews of RII comment about the hotel. So, Kin visits tripadvisor.com, which is the most popular website for tourist services. At tripadvisor.com, RII has a review pool RP containing 3,614 reviews (Figure 5). Kin does not have time to read all of them. Besides, she wants to read only the reviews that give opinions on her need: whether the bed is comfortable, whether the room is clean, whether the staffs are polite, whether the internet connection is good. The size of 3,614 reviews overwhelms Kin when she attempts to identify the helpful reviews by skimming the review webpages manually. Thus, Kin goes to the USRAnalyzer for help. Hereinafter shows how the USRAnalyzer helps Kin find helpful reviews satisfying her need.

When Kin opens the USRAnalyzer, its Helpful Review Finder welcomes Kin and prompts her to input her need (Figure 6). Kin describes it in her own language and style. In Figure 6, we show the USRAnalyzer's prompt and Kin's input c_j . When she submits it, the USRAnalyzer immediately starts to process c_j behind scene. The task is to identify the key-phrases in Kin's input (refer to next paragraph).



Rosen Inn International has received a large number (3,614) of online reviews at tripadvisor.com.

Figure 5. A Large Review Pool at tripadvisor.com

The key-phrases express Kin’s need whereas none key-phrases do not. For example, in Figure 6, the phrases such as “bet is comfortable” and “room is clean” are key-phrases expressing Kin’s need. But, the phrases “I want to find” and “I am interested in” are none key-phrases since they tell a little about Kin’s need. Then, the USRAalyzer expands the key-phrases with synonyms and acronyms extracted from a lexicon base, which can be an external lexicon repository or internal lexicon repository. An external lexicon repository is usually freely available on the Internet, for instance, WordNet or thesaurus.com. An internal lexicon repository is generally provided by the review website internally. As an example, we have built a lexicon repository that contains 132,761 English synonyms and acronyms from WordNet and thesaurus.com. The internal lexicon base supports our evaluation experiments of the USRAalyzer (see Chapter 7). We will elaborate the input processing in section 4.3.

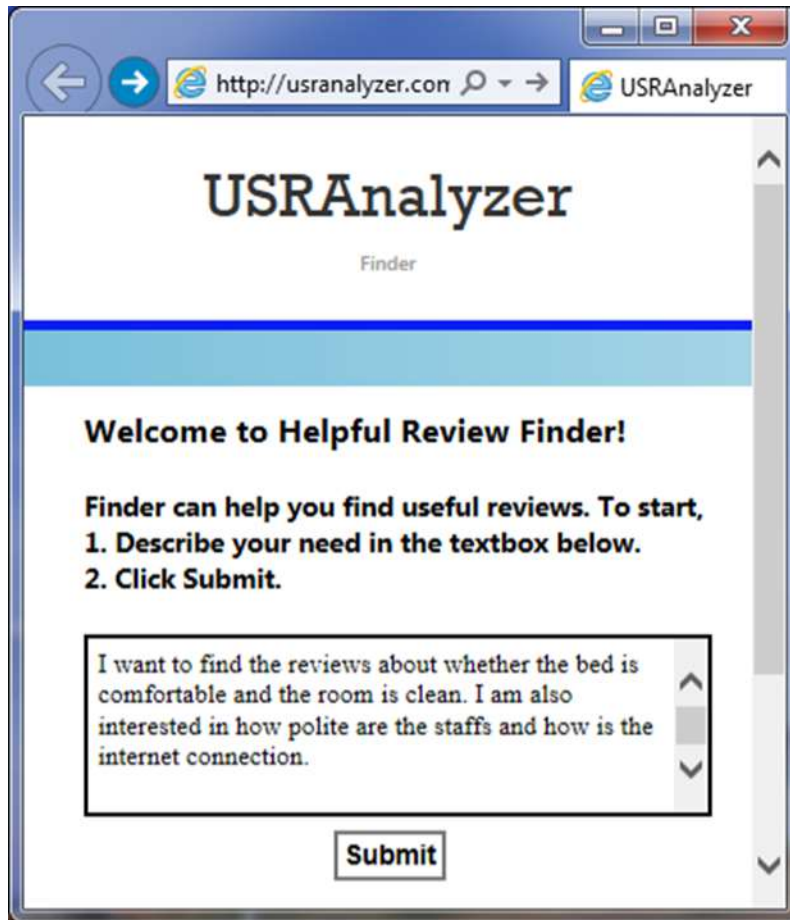


Figure 6 . An Example of Consumer Inputs

After processing c_j , the USRAnalyzer continues working behind scene. The task is to perform a preliminary evaluation of the relevance of each review in the review pool RP containing 3,614 RII reviews. This task is shown by box 3 of Figure 4. The evaluation can be either a ranking approach or a non-ranking approach, which we discussed in section 2.2.2 (chapter 2). For the effectiveness, we design the USRAnalyzer using the non-ranking approach that section 4.3.3 and section 5.1 will provide detailed description.

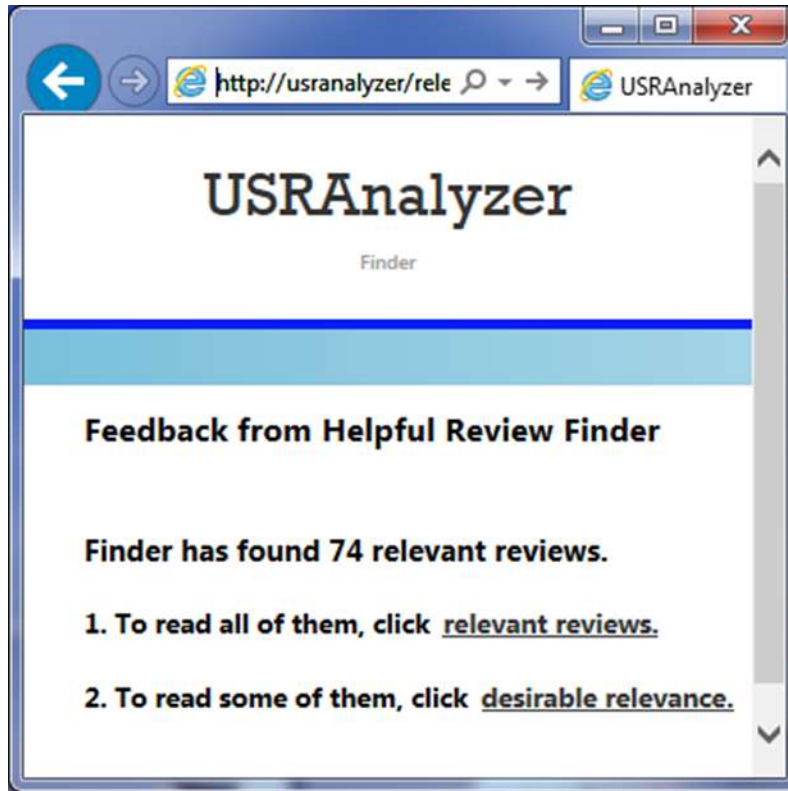


Figure 7. An Example of Feedback to Consumer

After the USRAnalyzer identifies the relevant reviews, the artifact provides feedback to Kin (Figure 7). The Helpful Review Finder informs Kin that there are 74 reviews relevant to her input (box 4 of Figure 4). The Helpful Review Finder also allows Kin to choose between reading all the relevant reviews and providing a desirable relevance level (d_j). In this case, we suppose that Kin chooses to select the d_j , and therefore, clicks the [desirable relevance](#) link. After Kin clicks it, the USRAnalyzer allows her to select d_j . In Figure 8, we show that Kin submits the very high relevance level, the highest level.

As soon as Kin submits her selection, the USRAnalyzer starts to assess, behind scene, the similarity between each relevant review and Kin's input. The assessment can use one of the

popular similarity measure used in web personalization research, including Jacard, cosine, etc. For accurate similarity assessment, we extend the similarity measure proposed by Sahami & Heilman (2006). We will present our proposed the similarity measure in section 5.2.

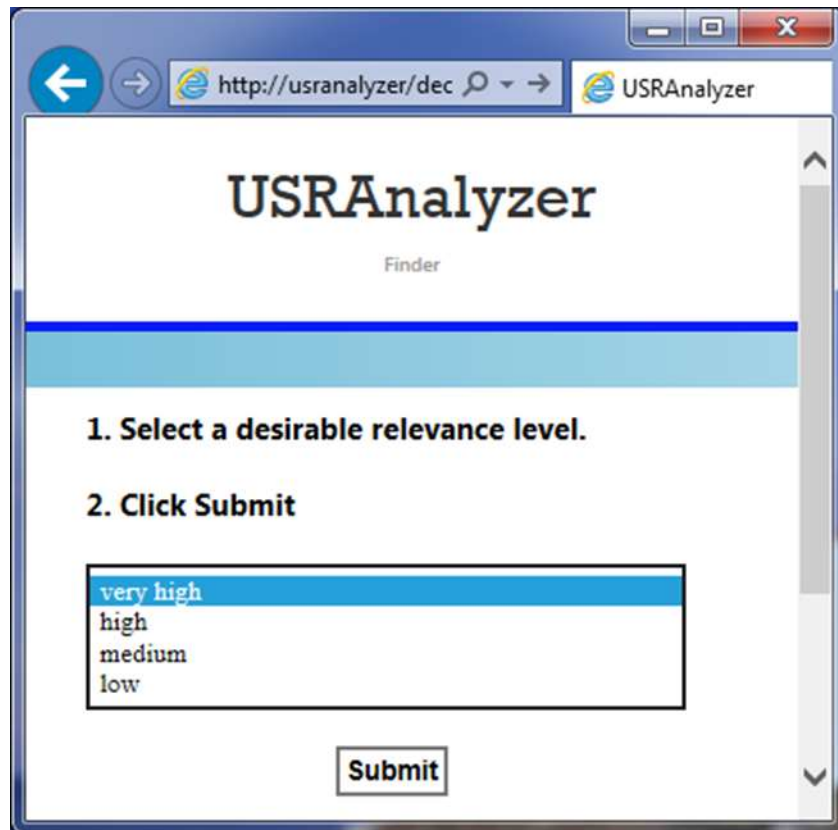


Figure 8. An Example of Desirable Relevance Level

Using the proposed similarity measure, the USRAnalyzer ranks the reviews by the similarity and sorts them in the descending order. Then, the USRAnalyzer displays to K_{in} those reviews with a relevance level l_j meeting $l_j \geq d_j$ (box 5 of Figure 4). In this use case, there are five reviews with a very high relevance level. We present the final result in Figure 9.

After Kin reads the selected five reviews, she may be satisfied and close the USRAnalyzer. Alternatively, she may want to get additional reviews. Then, she can run the USRAnalyzer again by clicking the Finder link below the name of the USRAnalyzer (see Figure 9). The interaction between it and Kin will repeat the steps illustrated above. In this way, Kin can use the USRAnalyzer as long as she wants.

As a final note, the review texts displayed in Figure 9 are downloaded from tripadvisor.com. We did not download the names of the reviewers since they are not the major issue for the USRAnalyzer. We assigned an automatic number (e.g. 5#) to each review according to its order in the downloading. Thus, the number of a review in the final result identifies the review. In Figure 9, the readers can see that the USRAnalyzer presents the final desirable reviews by their relevance levels rather than their stored orders. Thus, this use case provides the strong evidence that supports the USRAnalyzer's efficacy in offering dynamic personalized utility-sensitive analysis (DPUSRA).

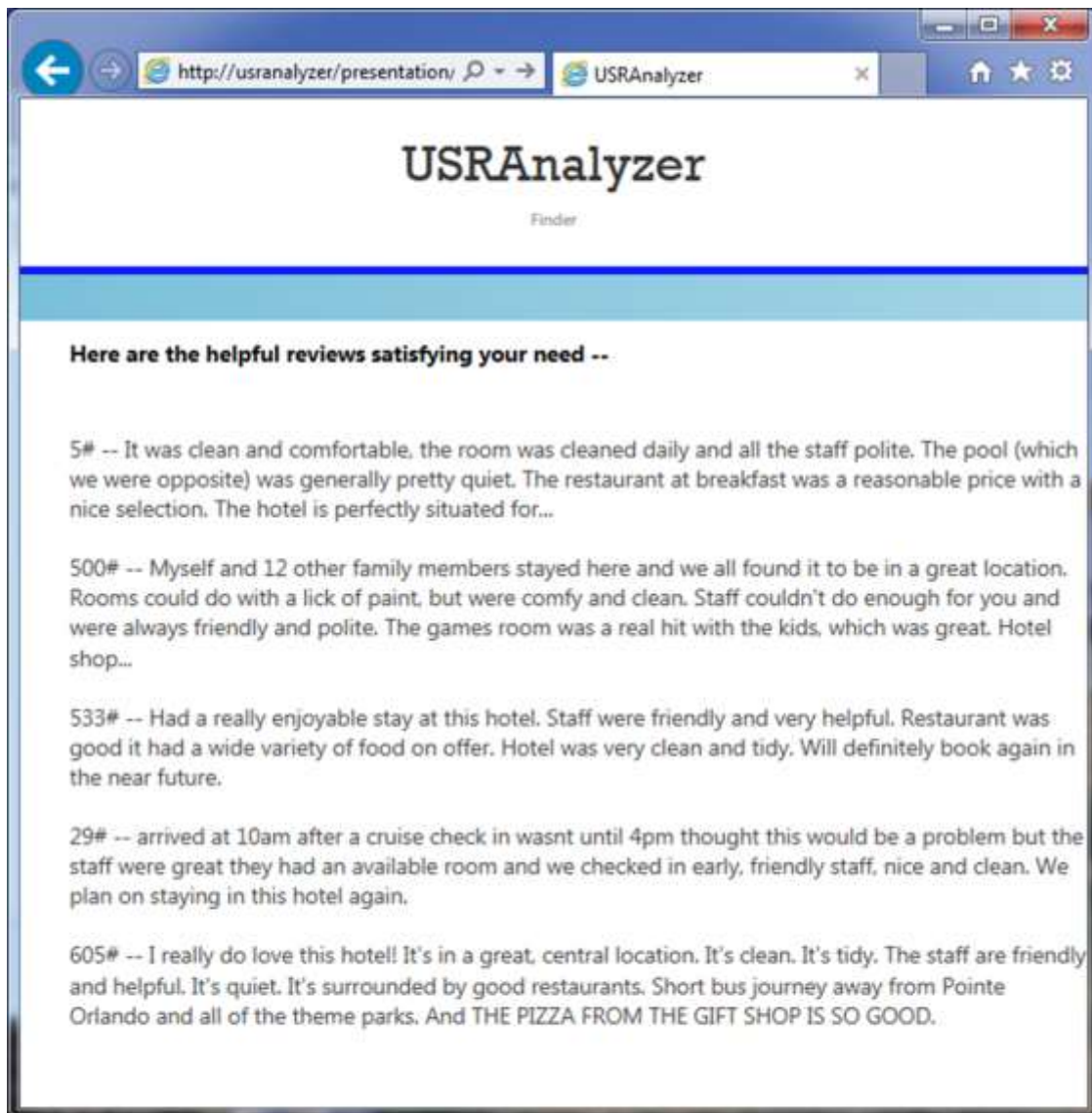


Figure 9. An Example of Resultant Helpful Reviews

4.3 USRAnalyzer Architecture

In this section, we present our proposed general-purpose architecture for the USRAnalyzer, which aligns with its overview model we presented in section 4.1. In Figure 10, we outline the USRAnalyzer architecture including Graphic User Interface (GUI, front end) and Back End. The GUI includes three modules (module 1, 4, and 6) which are User Search Initiation, User Relevance Criterion, and Helpful Review Presentation. The Back End also consists of three modules (module 2, 3, 5) which are Input Processing, Relevance Evaluation, and Review Selection. So, there are totally six modules functioning seamlessly to interact with consumers and provide personalized helpful reviews. Each module contains one or two component(s). In following sub-sections, we describe each module in detail.

The readers may notice that in Figure 10, we shade the Language Translator component of module 1, Input Processing. This is because our prototype has not implemented that component for this reason: We want to evaluate the core functionality of the USRAnalyzer in our experiments described in chapter 7. The Language Translator component is a useful but not a core function of the USRAnalyzer. To facilitate website designers to apply the USRAnalyzer, we have provided it with the capability to work with many software tools that provide advanced language translation. A very useful tool of language translation is Google Search. The USRAnalyzer can easily work with the language translation tools to accomplish the task of the Language Translator.

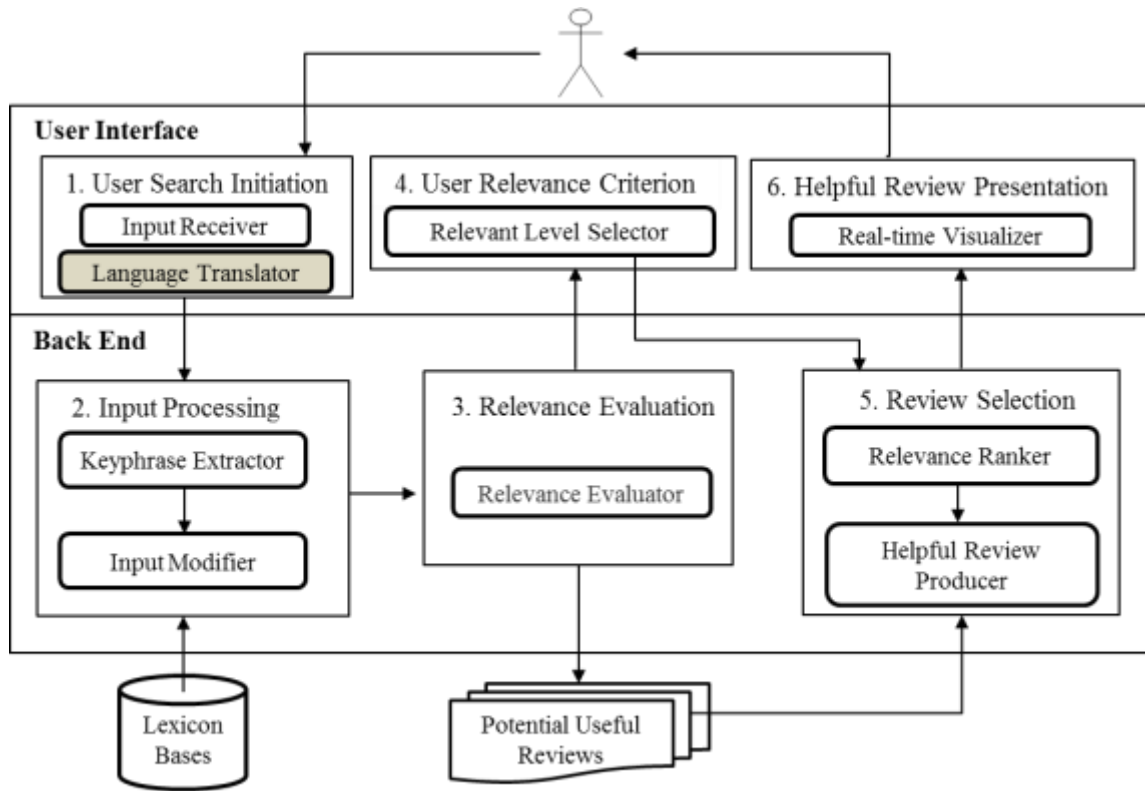


Figure 10. USRAnalyzer Architecture

4.3.1 User Search Initiation

A user starts a search for helpful reviews at the Input Receiver of the User Search Initiation module. Here, the Input Receiver prompts the consumer to input his/her need. This is regarded as the most effective way for the USRAnalyzer to execute its tasks accurately (Fasolo et al., 2005; Wang & Benbasat, 2009; Reisen & Hoffrage, 2010). Also, the Input Receiver allows the consumer to input the need in her/his own style without imposing any restriction. Such flexibility is a key for effective consumer decision support (Atahan & Sarkar, 2011; Armentano et al., 2006; Song et al., 2007).

When the Input Receiver receives the user's input, the Language Translator will run if the input is in a natural language different from the natural language used by the reviews. In such case, the User Search Initiation module will pass the translated user input to the Input Processing module (see section 4.3.2). If there is not a need for language translation, the Language Translator will not run. And, the User Search Initiation module will pass the user input directly to the Input Processing module.

4.3.2 Input Processing

The Input Processing module runs when it receives the user input or the translated user input from the User Search Initiation module. The Keyphrase Extractor runs to extract the key phrases from the user input. A key phrase is typically a noun phrase that expresses the user need. But, a phrase in the user input may not be a key phrase (Baroni & Lenci, 2010). Take this user input as example, "I want to find the reviews about how good the Internet connection is at the hotel." The phrase "I want to find" tells little about the user need. But, the phrase "how good the Internet connection is" expresses that the user need; that is, the user wants the hotel provides good Internet connection. A useful review should be relevant to that user need. So, key-phrase extraction is critical for the USRAnalyzer to personalize helpful reviews. Web personalization (WP) research has shown that a WP artifact needs to extract the noun phrases from the user input. They together express the need of the user (Baroni & Lenci, 2010; Song et al., 2007; Zettlemoyer & Collins, 2009). The types of noun phrases include adjective+noun (e.g. friendly staffs), noun+verb (e.g. Internet connect), noun+verb+adjective|noun (e.g. hotel provides good

WiFi). The USRAnalyzer can extract all those noun phrases using the common noun-phrase extraction approach used by WP research.

Many WP researches have proposed methods to identify noun phrases in user inputs. The methods commonly involve part-of-speech (POS) analysis of the user input. Then, the noun-phrases are extracted according to the analysis (Haddoud & Abdeddaim, 2014). Many text processing tools include the POS analysis; the examples include IBM Watson and SAS Text Miner. So far, the common POS approach implemented in those tools represents advanced implementation of POS technology. The Keyphrase Extractor uses the common POS approach utilized in natural language processing to accomplish the none-phrase extraction task. The extracted none-phrases will be used by the Relevance Evaluation and the Review Selection (section 4.3.3 and section 4.3.4).

After the Keyphrase Extractor extracts the noun phrases, the Input Modifier extends the key phrases in order to produce the appropriate coverage of relevant reviews. Particularly, the Input Modifier extends the noun phrases with synonyms and acronyms since many reviews may not use the exact words used in the user input. Instead, the reviews use the synonyms and acronyms in the noun phrases. Therefore, like other web personalization artifacts, the USRAnalyzer needs to identify the reviews that use synonyms and acronyms in the noun phrases because those reviews should be relevant to the user input (Cao et al., 2007; Wang et al., 2014). The Input Modifier extends the noun phrases with lexicon-based input processing approach for the reason presented as following.

In section 2.1, we discussed input processing techniques of three types: model-based, feedback-based, and lexicon-based. The lexicon-based is the best for the USRAnalyzer because model-based approach requires predefined consumer models and is suitable for record-based personalization (Rattenbury & Naaman, 2009; Rendle, 2012). Yet, the USRAnalyzer is an interactive web personalization artifact. Its users are often first-time visitors. They have not conducted the past visits on which the user models are derived. For similar reason, the feedback-based approach is not suitable for the USRAnalyzer. Additionally, several researches have suggested that feedback-based approach tends to impose burden on users. Since feedback-based approach typically requires the users refine their inputs, the users may feel burdensome to do so (Rendle, 2012; Song et al., 2007; Wang & Benbasat, 2009; Wang et al., 2014). Therefore, we employ lexicon-based input processing. The Input Modifier can work with freely-available online lexicon repositories, e.g. thesaurus.com and WordNet. Moreover, the Input Modifier can work with multiple natural languages (e.g. English, Japanese, Chinese, Spanish, and German) when the Language Translator is implemented. So, the USRAnalyzer is a multi-lingual web personalization artifact and has broad practical utility (Ambati & Uppuluri, 2006; Cao et al., 2007).

4.3.3 Relevance Evaluation

When the Input Modifier finishes input processing, the Relevance Evaluator conducts the initial assessment on the relevance of each review in the review pool. The objective is to identify the set of relevant reviews or the relevant set. In other words, the Relevance Evaluator performs only the initial relevance assessment out of the two relevance assessments performed by the

USRAnalyzer. It performs relevance evaluation twice (two rounds), and the Relevance Evaluator performs the first round. Two-round evaluation can significantly improve the accuracy of the output helpful reviews. Often, a single evaluation produces large errors. And, the effective way to reduce them is to perform relevance evaluation multiple times. Smyth & Balfe (2006) demonstrated that a web personalization system can achieve much higher accuracy by performing relevance evaluation twice. The second evaluation can significantly improve the result from the first evaluation, and thus, raise evaluation accuracy. Other researches have drawn similar conclusion (e.g. Leveling & Jones, 2010; Ogilvie et al., 2009). Thus, we propose that the USRAnalyzer performs relevance evaluation twice. The Relevance Evaluator performs the initial evaluation. Then, the Relevance Ranker (see section 4.3.4) performs the second assessment. The initial evaluation executes our proposed DPSO-KM algorithm (see section 5.1). Here we provide a brief discussion the Directed Particle Swarm Optimization and K-Means (DPSO-KM) algorithm that extends the prior PSO-KM algorithm proposed by web personalization research.

Web personalization research uses dozen approaches to identify the relevant reviews. Main approaches are clustering and Support Vector Machine (SVM). SVM is more appropriate to evaluate the relevance of long texts with more than 200 words. Clustering not only works well for long texts, but also is particularly suitable to relevance evaluation of short texts (less than 200 words) (Cagnina et al., 2014; Labroche et al., 2003). Recently, the Particle Swarm Optimization and K-Means (PSO-KM) approach has been proven high effective and efficiency for relevance evaluation of short texts (Cagnina et al., 2014; Cui & Potok, 2005). The PSO-KM is a global optimization method for webpage relevance evaluation. The method first performs a global

search to determine the initial centroids. Then, K-Mean uses them to iteratively produce the final clusters. In our current research, the Relevance Evaluator evaluates online reviews that mix predominant short texts with a small number of long texts. Thus, the PSO-KM approach with appropriate improvement is very suitable for the Relevance Evaluator. In section 5.1, we will present our proposed Directed Particle Swarm Optimization and K-Means (DPSO-KM) algorithm that improves the PSO-KM approach. The Relevance Evaluator runs the DPSO-KM to produce the relevant reviews, which are then stored in the Potential Useful Reviews storage (Figure 10).

4.3.4 User Relevance Criterion, Review Selection, Helpful Review Presentation

The User Relevance Criterion module (module 4 in Figure 10) runs when the relevant reviews have stored by the Relevance Evaluator. The Relevant Level Selector interfaces with the user after the Relevance Evaluator produces the relevant set. The Relevant Level Selector allows the user to select the desirable relevance level. A user-determined desirable relevance can more accurately represent the user's need than a system-determined desirable relevance. The latter approach often imposes a stiff criterion that is not suitable to individual users (Atahan & Sarkar, 2011; Micarelli & Sciarrone, 2004). That is the reason why the Relevant Level Selector allows the user to select the desirable relevance level from the list of relevance levels such as very high, high, medium, and low. Such relevance levels can adapt a wide range of desirable relevance options.

When the user submits the desirable relevance level, the Relevant Level Selector passes it to the Relevant Ranker in the Review Selection module (module 5). The Relevance Ranker runs our proposed Review Utility Ranking (RURanking) algorithm (see section 5.2) to produce the ranked list of the relevant reviews with the most helpful reviews at the top of the list.

When the Relevant Ranker outputs the desirable helpful reviews, the Helpful Review Producer in module 5 extracts the reviews from the Potential Useful Reviews storage that stores the relevant reviews produced by the Relevance Evaluator in module 3 (see section 4.3.3). Then, the Helpful Review Producer passes the extracted reviews to the Helpful Review Presentation module (module 6) where the Real-time Visualizer presents the user with the review outputs. Notably, all final helpful reviews need to meet the desirable relevance level criterion; that is, the relevance level of any helpful review output is greater than or equal to the desirable relevance level (d_j).

CHAPTER 5 ALGORITHMS

In section 4.3.3, we presented that the Relevance Evaluation module in the USRAnalyzer architecture runs our proposed algorithm of the Directed Particle Swarm Optimization and K-Means (DPSO-KM) to accomplish the initial relevance assessment. In section 4.3.4, we discussed that the Review Selection module runs our proposed Review Utility Ranking (RURanking) algorithm to rank the relevant reviews by relevance level. In this chapter, we present the two algorithms, the DPSO-KM algorithm in section 5.1 and the RURanking algorithm in section 5.2.

5.1 Directed Particle Swarm Optimization and K-Means Algorithm

We have proposed the Directed Particle Swarm Optimization and K-Means (DPSO-KM) algorithm that is an extension of the Particle Swarm Optimization and K-Means (PSO-KM) method proposed by prior research. In this section, we discuss PSO-KM method in sub-section 5.1.1, and present our proposed DPSO-KM algorithm in sub-section 5.1.2. Also, we provide an overview of K-Means clustering in Appendix I.

5.1.1 Prior PSO-KM Method

In section 4.3.3, we introduced prior PSO-KM method. Several PSO-KM proposals appeared in prior research. Although different proposals are different in the heuristics used to perform the global search known as Particle Swarm Optimization (PSO), all the proposals follow

the same general procedure. In this sub-section, we discuss the common PSO-KM procedure on the basis of Cagnina et al. (2014) and Cui & Potok (2005).

The PSO-KM method includes two main procedures, Particle Swarm Optimization (PSO) and K-Means (KM). The PSO procedure is a stochastic optimization algorithm performing global search to identify the centroids for KM procedure. The PSO uses the cluster vector space (CVS) where a potential clustering solution is a cluster vector. Typically, the vector consists of the terms identified by the key-phrase selection (refer to section 4.3.2 for detail). Most time, the terms are weighed, and the weights are calculated by TF-IDF algorithm:

$$w_{hg} = f_{t,hg} * \log \frac{N}{f_{p,hg}} \quad (F1)$$

In F1, w_{hg} is the weight of term g ; $f_{t,hg}$ is the frequency of term g in webpage h ; $f_{p,hg}$ is the frequency of term g in the webpage collection; N is the number of the webpages in the collection.

The POS global search is an iterative process designed to achieve the best fitness value of the cluster vectors each of which is evaluated by some validation measures known as Internal Clustering Validity Measures (ICVMs) (Cagnina et al., 2014). The ICVMs are a set of measures that can provide statistical validation of the results produced by the global search. The validation is performed in every iteration cycle during which a cluster vector represents a position in the CVS. The position must meet the requirement of the global best position called swarm and the

local best position called particle. In a subsequent iteration cycle, the particles move according to the updating functions:

$$v_{id} = w (v_{id} + r_1 (l_d - p_{id}) + r_2 (s_d - p_{id})) \quad (\text{F2})$$

$$p_{id} = p_{id} + v_{id} \quad (\text{F3})$$

In F2 and F3, v_{id} is the velocity of particle i at the dimension d ; w is the inertia factor to balance particle and swarm; r_1 is a so-called personal learning factor to ensure optimal particle; r_2 is a so-called social learning factor to ensure optimal swarm; p_{id} is the position of the particle i at the dimension d ; l_d is the particle at the dimension d ; s_d is the swarm at the dimension d . In the literature, there are different versions of F2 while F3 is commonly used. The difference between two F2 versions generally comes from the difference between the corresponding PSO heuristics.

The goal of the PSO is to ensure global optimization of the K-Means (KM) procedure which is performed by using the centroids produced by the PSO. The researches generally use the KM procedure as implemented by data mining and statistic software packages (e.g. SAS and SPSS). The KM usually employs the Euclidean distance for similarity measure (Cagnina et al., 2014; Cui & Potok, 2005). In Appendix I, we give a detailed discussion on K-Means clustering. It is the most popular clustering technique with very effective performance.

According to Cagnina et al. (2014), the PSO-KM is the most accurate algorithm for webpage relevance evaluation. However, the PSO tends to be computationally expensive when the webpage collection is large. Thus, the researchers have attempted to improve the efficiency

of the PSO (Cagnina et al., 2014; Cui & Potok, 2005). Although Cagnina et al. (2014) proposed, according to the researchers, the most efficient PSO, their proposal still involved complex computation. Because utility-sensitive review analysis (USRA) is usually performed on a large review pool, the PSO cannot meet the efficiency requirement. Therefore, we propose the DPSO-KM algorithm (see section 5.1.2) to eliminate the inefficiency of the PSO. At the same time, the DPSO-KM algorithm preserves the effectiveness of the PSO-KM method.

5.1.2 Proposed DPSO-KM Algorithm

In section 5.1.1, we discussed that the PSO-KM is the most effective method for identifying relevant reviews from a large review pool. However, the PSO procedure tends to be inefficient, and ongoing research effort is attempted to improve the efficiency. To eliminate the inefficiency of the current PSO method, we propose the Directed Particle Swarm Optimization and K-Means (DPSO-KM) algorithm. Indicated by its name, the DPSO-KM improves the PSO by our proposed DPSO that utilizes the left join of online review r_g and expanded consumer input c . In other words, our proposed DPSO-KM does not perform global search, and thus, does not use F1, F2, and F3 (refer to section 5.1.1).

Also, our DPSO-KM utilizes the KM method G-means proposed by Hamerly & Elkan (2004). Their method iteratively identifies the optimal centroids by using initial centroids. In each KM iteration cycle, the G-means method replaces some initial centroid with two new centroids when the cluster is not approximate to the Gaussian criterion commonly used in statistical analysis. Such centroid replacement will iterate until each resultant cluster is

approximate to the Gaussian criterion. The algorithm of G-means is presented in Figure 11, which comes from the Algorithm 1 provided by Hamerly & Elkan (2004). We use the G-means for two main reasons. First, the G-means is widely used in information retrieval and has proven to be very effective (Manning et al., 2008). Second, the G-means can help ensure high accuracy of the output of relevant reviews, i.e., relevant set RS. This advantage of the G-means is supported not only by Manning et al. (2008), but also by our trail-and-error design process. Our design experience have shown that the G-means KM method can achieve higher accuracy of review relevance evaluation than support vector machine (SVM) method and other KM methods (e.g. MacQueen algorithm and Lloyd-Forgy algorithm; refer to Appendix I of this paper).

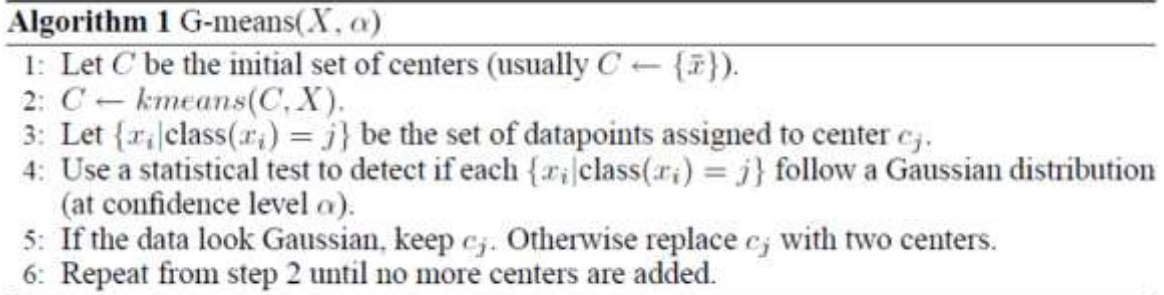


Figure 11. G-means from Hamerly & Elkan (2004)

In Figure 12, we present the heuristic of our proposed DPSO-KM starting with retrieving the review pool RP stored in the review repository of the review website. Then, the algorithm uses r_g ($r_g \in RP$) to represent a review in RP, c to represent the expanded user input, and χ_L to represent the left join operator. Also, the DPSO-KM use J for the set of initial cluster centroids used by the K-Means procedure, a_g ($a_g \in J$) for an initial centroid, Ψ for a set of clusters produced

by the KM procedure, M_h ($M_h \in \Psi$) for the cluster h generated by the KM, and RS for the relevant set.

The main procedure of the DPSO-KM starts at Line 1 where the set of initial centroids is empty since the KM has not run yet. From Line 2 through Line 7, the DPSO-KM loops through each review in RP to evaluate the intersection of each review g and the expanded consumer input c (Line 3). In general, if the intersection takes a large portion of c , the review g is highly relevant to the user input. This ensures the left join a_g (Line 4) to be highly relevant to the user input, and thus, meet the swarm criterion of the prior PSO-KM method (refer to section 5.1.1). To ensure high relevance, the DPSO-KM uses to produce the left joins, only reviews that have a large intersection with c ; that is, $|r_g \cap c| / |c| > s$ (the significant value). The s can be given the value 0.1 according to the common practice of the text mining (Berry & Linoff, 2010). If the intersection takes a larger portion greater than s (Line 3), the DPSO-KM outputs the left join a_g of r_g and c (Line 4). Then, a_g is added to the set (J) of left joins (Line 5). The following example, Example I, illustrates the loop from Line 2 through Line 7.

Example I:

Suppose the expanded user input c and the review r_5 ($r_5 \in RP$) are as following:

$c =$ “friendly/courteous staffs/employees, comfortable/comfy beds”

$r_5 =$ “I was welcomed by friendly staffs and clean rooms.”

Then, since the intersection of r_5 and c contains only one phrase, i.e. “friendly staffs,” yet c contains two phrases separated by the comma. Therefore,

$$|r_5 \cap c|/|c| = |\text{friendly staffs}|/|c| = 1/2 = 0.5 > 0.1$$

Because the intersection takes a larger portion than 0.1, the DPSO-KM produces the left join ($a_5 \leftarrow r_5 \chi_L c$) as the following,

$a_5 =$ “I was welcomed by friendly staffs and clean rooms.”

Then, the a_5 given above is added to the set J that will be used as the set of initial centroids in the following K-Means procedure.

We shall stress that in Line 5, the DPSO-KM produces the J containing all the left outer joins meeting the criterion give in Line 6. Then, the algorithm uses J as the set of initial centroids for K-Means (KM) that starts at Line 8. The KM is an iterative process performed in Line 8. The readers may refer to Figure 11, the G-means algorithm, which our KM procedure executes.

Specifically, in Line 8, the DPSO-KM performs the KM iterative procedure using the review pool RP and J as the set of initial centroids. In the first iteration cycle, the KM procedure produces a set of initial clusters each of which centers an initial centroid a_g . Then, each initial cluster is evaluated for whether it is approximate to the Gaussian criterion. If an initial cluster is approximate to the Gaussian criterion, then that initial centroid is kept. If an initial cluster is not approximate to the Gaussian criterion, then two new centroids will be found.

LET

RP be initially generated Set of Reviews (Review Pool)

r_g ($r_g \in RP$) be individual review g

c be expanded consumer input

χ_L be Left Outer Join operator

J be set of initial Cluster Centroids to be used by the K-Means algorithm

a_g ($a_g \in J$) be initial Cluster Centroid g

Ψ be set of Clusters returned by the K-Means algorithm

M_h ($M_h \in \Psi$) be a cluster "h" of Reviews generated by the K-Means

RS be pruned set of reviews

1. $J \leftarrow \Phi$

2. **For** $1 \leq g \leq |RP|$ **DO**

3. **IF** $|r_g \cap c| / |c| > s$

4. **THEN** $a_g \leftarrow r_g \chi_L c$

5. $J \leftarrow J \cup a_g$

6. **END IF**

7. **END FOR**

8. $\Psi \leftarrow \text{K-Means}(RP, J)$

9. $RS \leftarrow \emptyset$

10. **FOR** each $M_h \in \Psi$ **DO**

11. **IF** $M_h \cap J \neq \emptyset$

12. **THEN** $RS \leftarrow RS \cup M_h$

13. **ENDIF**

14. $\Psi \leftarrow \Psi - M_h$

15. **ENDFOR**

Figure 12. DPSO-KMPseudo Code

Then, the second KM iteration cycle is executed using the new set of centroids that was produced by the first iteration. The second KM iteration cycle repeats the same procedure as used in the first iteration cycle to produce a new set of clusters. Then, each cluster produced by the second iteration cycle is evaluated for whether it is approximate to the Gaussian criterion. If there is still some centroid that does not meet the criterion, the KM starts the third iteration cycle. At the end of the third KM iteration cycle, if there is still some centroid that does not meet the criterion, the fourth KM iteration cycle will be started.

In the same fashion described above, each new KM iteration cycle refines the clusters produced by the previous iteration cycle. At the end of the new iteration cycle, each cluster is evaluated for whether it is approximate to the Gaussian criterion. If a cluster is not approximate to the Gaussian criterion, then two new centroids are found. The KM iteration cycle will continue until ending at a set of review clusters each of which is approximate to the Gaussian criterion.

When the KM iteration ends, it produces a set (Ψ) of review clusters (refer to Line 8). Some resultant clusters contain left join $a_g \in J$ while other clusters do not contain it. Intuitively, if a review cluster M_h (Line 11) contains the left join, the reviews in that cluster are relevant to the user input. Thus, if a review cluster M_h contains the left join $a_g \in J$, M_h will be added to the relevant set RS (Line 12).

In contrast, other KM-output clusters do not contain left join a_g ($a_g \in J$). Those clusters consist of the reviews that are irrelevant to the user. Thus, the DPSO-KM rejects those clusters

(Line 14). Finally, at the end of the loop from Line 10 to Line 14, the DPSO-KM outputs the relevant set RS that will be used by our proposed RURanking algorithm described in section 5.2.

To illustrate the capability of our proposed DPSO-KM algorithm, we provide a simple example, Example II, next.

Example II:

Suppose the expanded user input c and the four reviews in the review pool RP are as following:

$c = \text{“friendly/courteous staffs/employees, comfortable/comfy beds”}$

$r_5 = \text{“I was welcomed by friendly staffs and clean rooms.”}$

$r_{37} = \text{“The beds are too small, so you may not get rest.”}$

$r_{218} = \text{“There are specious rooms and quite environment.”}$

$r_{632} = \text{“We enjoyed comfortable beds and courteous employees.”}$

Therefore, the intersection of each review (r_5 , r_{37} , r_{218} , and r_{632}) and c is as following:

$$|r_5 \cap c|/|c| = 1/2 = 0.5$$

$$|r_{37} \cap c|/|c| = 0/2 = 0$$

$$|r_{218} \cap c|/|c| = 0/2 = 0$$

$$|r_{632} \cap c|/|c| = 2/2=1$$

Notably, the second and the third intersections are empty because no phrase of the expanded user input c is included in r_{37} and r_{218} . Thus, the DPSO-KM produces two left joins, $a_5 \leftarrow r_5 \chi_L c$ and $a_{632} \leftarrow r_{632} \chi_L c$ as following,

$a_5 =$ “I was welcomed by friendly staffs and clean rooms.”

$a_{632} =$ “We enjoyed comfortable beds and courteous employees.”

Then, the set (J) of initial centroids for the KM consists of a_5 and a_{632} :

$$J = \{a_5, a_{632}\}$$

Next, the KM iterative process starts. In the first interaction cycle, the KM produces two clusters supposed to look like Cluster (1) and Cluster (2), which are presented in Figure 13. Supposing also that Cluster (1) is approximate to the Gaussian criterion and that Cluster(2) is not. Thus, the DPSO-KM keeps a_{632} while finding two new centroids.

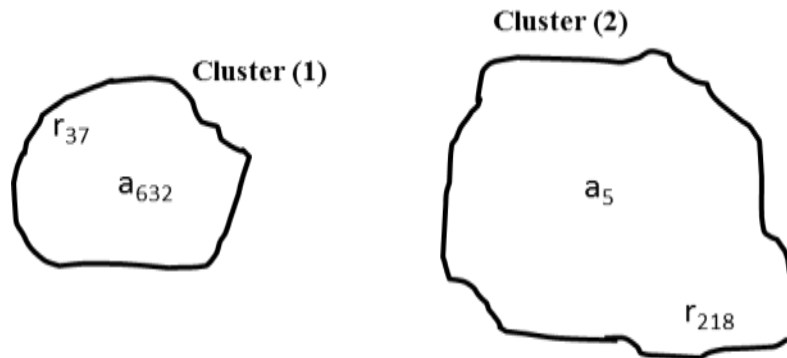


Figure 13. Example II, Two First-Cycle Clusters

Let's suppose the new centroids are a_{11} ($a_{11} \in RP$) and a_{953} ($a_{953} \in RP$). Then, the DPSO-KM performs the second iteration cycle using a_{632} , a_{11} , a_{953} as the centroids. Let's also suppose that the second KM iteration outputs three clusters as shown in Figure 14. And, they all approximate the Gaussian criterion. Thus, the KM iteration ends when the three clusters are output.

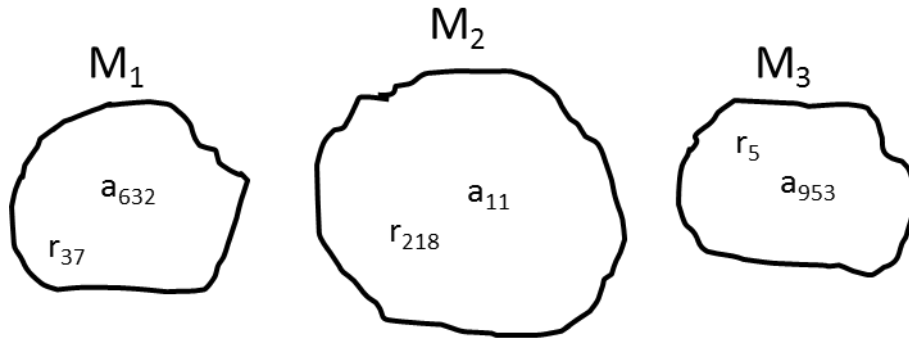


Figure 14. Example II, Final KM Clusters

Let's suppose the three resultant clusters are M_1 , M_2 , and M_3 as depicted in Figure 14. Since M_2 does not contain left join (i.e. a_5 or a_{632}), the DPSO-KM rejects M_2 . However, M_1 contains left join a_{632} and M_3 contains left join a_5 ($a_5 = r_5$) Thus, the DPSO-KM adds M_1 and M_3 to get the set of relevant review (RS). The output RS may look like Figure 15:

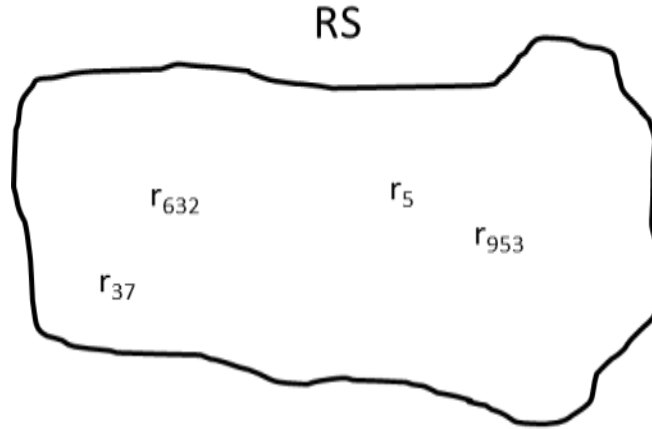


Figure 15. Example II, Relevant Set (RS)

In summary, given the description and examples above, we can see that the final output RS can be seen as a maximum set of reviews relevant to the user input. And, the reviews in the RS meeting PSO-KM's requirement of the global best relevant set called swarm and the local best relevant set called particle (refer to section 5.1.1). This is because the RS is highly relevant to the expanded user input c while the output the Ψ does not. Consequently, the RS amounts to the final PSO-KM output that is optimal.

The DPSO-KM heuristic has three advantages compared to the PSO-KM. First, it eliminates the global search on which the latter relies. Thus, our proposed DPSO-KM is more efficient than the PSO-KM since the PSO's global search is inefficient (see section 5.1.1). Second, the DPSO-KM does not use predefined threshold, eliminating the need for threshold configuration. Thus, the DPSO-KM is much easier to use than the PSO-KM since threshold configuration requires extensive testing and complex heuristics. Third, the previous advantages of the DPSO-KM make its implementation and maintenance simpler comparing to the PSO-KM.

5.2 Review Utility Ranking Algorithm

In section 4.3.4, we mentioned that the Relevance Ranker runs our proposed Review Utility Ranking (RURanking) algorithm to rank the relevant reviews by relevance level. In this section, we present the Review Utility Ranking (RURanking) algorithm to rank each review in the relevant set RS output by our DPSO-KM algorithm (see section 5.1.2). The RURanking algorithm evaluates the similarity between relevant review and the expanded consumer input. In general, the RURanking can utilize any similarity measure used by web personalization research. For example, many researches use Jaccard index or cosine similarity (Baeza-Yates & Ribeiro-Neto, 2011; Kopliku et al., 2014; Yin et al., 2009). But, similarity measures for short text (less than 200 words) is superior over the other similarity measures in our case since consumer inputs and majority online reviews tend to be short text. Therefore, we consider the popular similarity measure of the Web-based Similarity Kernel (WSK) contributed by Sahami & Heilman (2006). Its Function is as F4:

$$Sim(p_g, r_j) = (pf_{g,j}) * \log \left[\frac{T}{rf_g} \right] \quad (F4)$$

In F4, p_g is g^{th} noun-phrase in the expanded user input; r_j is j^{th} relevant review; $pf_{g,j}$ is the frequency of noun-phrase p_g in review r_j ; T is the total number of the reviews in the review pool; rf_g is frequency of noun-phrase p_g in the review pool.

Although the WSK given by F4 is a very popular similarity measure for short texts, yet it is not adequate for our proposed RURanking algorithm. We previously discussed that online

reviews tend to be short texts. Still, some reviews exceed 200 words considered as long texts. Thus, we need to extend the WSK so that the RURanking can appropriately evaluate the similarity between user input and long review.

Therefore, we propose an extension of WSK and name the extension as ExWSK (see following formula ExWSK) which includes the same parameters as WSK (F4). However, the WSK does not include $pf_{g,j}$ in the log operation whereas ExWSK includes $pf_{g,j}$ in its log operation. The ExWSK can thus accommodate the long reviews. This is because log operation is less sensitive to the redundant counts of the noun-phrases in the reviews (Hamilton, 2012). When the reviews in a review pool are all short text, the redundant counts can be regarded as minimum. In such case, the WSK is adequate. But, a review pool may contain some long reviews that mix with short reviews. Since long review tends to increase the opportunity for repeating noun-phrases, the redundant counts of the noun-phrases cannot be regarded as minimum. They may inflate the $pf_{g,j}$ and the WSK. Yet, the redundant counts will have less influence on ExWSK since its log operation is less sensitive to the redundancy.

Extending the WSK, the ExWSK not only preserves the advantage of the WSK but also improves its capability of conducting proper similarity evaluation in the cases of short and long reviews. Consistent with the parameters of WSK, the ExWSK include parameter p_g as g^{th} noun-phrase in the expanded user input; r_j is j^{th} relevant review; $pf_{g,j}$ is the frequency of noun-phrase p_g in review r_j ; T is the total number of the reviews in the relevant set RS; rf_g is frequency of noun-phrase p_g in the relevant set RS.

$$Sim(p_g, r_j) = \log \left[pf_{g,j} \frac{T}{rf_g} \right] \quad (ExWSK)$$

In Figure 16, we present the heuristic of the RURanking. Line 01 through Line 03 counts T , $pf_{g,j}$, and rf_g respectively to find the frequencies used by the ExWSK. At Line 05, a loop starts to run ExWSK (Line 6) when rf_g is not zero. In Line 05 through Line 12, the loop adds into the desirable set DS, the review r_j when the similarity between r_j and p_g is greater than or equal to the cutoff point δ . Otherwise, the review is rejected. The cutoff point δ is in fact the desirable relevance level d_j selected by the consumer via the User Relevance Criterion interface of the USRAnalyzer. Thus, the final desirable set DS satisfies d_j selected by the user. The USRAnalyzer will present the user all the reviews in DS, which are sorted in the order of descending similarity to the use input (Line 13 and Line 14).

Unlike prior relevance ranking algorithm proposed by web personalization (WP) research, the RURanking algorithm utilizes the desirable relevance level d_j instead of a predefined threshold. Thus, the RURanking is consumer-centric and better satisfies consumer need. Furthermore, the RURanking is easier to use and maintain than the prior algorithm; the RURanking eliminates the need for complex configuration and maintenance of predefined threshold.

LET: p_g be g^{th} noun-phrase
 r_j be j^{th} relevant review
 $pf_{g,j}$ be the frequency of p_g in r_j
 rf_g be the frequency of p_g in RS
 T be number of the reviews in RS
 δ be cutoff point
 DS is a set of desirable reviews (desirable set)

- 01: $T \leftarrow \text{count } r_g$
- 02: $pf_{g,j} \leftarrow \text{count } p_g \text{ in } r_j$
- 03: $rf_g \leftarrow \text{count } p_g \text{ in } RS$
- 04: $DS \leftarrow \Phi$
- 05: **FOR** $rf_g \neq 0$ **DO**
- 06: $Sim(p_g, r_j) = \log\left[\frac{pf_{g,j}}{rf_g} \frac{T}{rf_g}\right]$ (*ExWSK*)
- 07: **IF** $Sim(p_g, r_j) \geq \delta$
- 08: **THEN** $DS \leftarrow DS \cup r_j$
- 09: **ELSE**
- 10: reject r_j
- 11: **ENDIF**
- 12: **ENDFOR**
- 13: **SORT** DS descending
- 14: **END**

Figure 16. RURanking Algorithm

CHAPTER 6 IMPLEMENTATION

In the previous chapters, we presented the USRAnalyzer's overview model (chapter 4), architecture (chapter 4), and two algorithms (chapter 5). This chapter introduces the implementation of the USRAnalyzer prototype, which is an instance of the proposed USRAnalyzer architecture (section 4.3). The Language Translator component has not been implemented in the prototype because the evaluations presented in Chapter 7 should focus on the main functions of the USRAnalyzer. In Figure 10 (see section 4.3), we presented that the USRAnalyzer consists of two major functional blocks, user interface (front end) and back end. We present their implementations in section 6.1 (user interface) and section 6.2 (back end).

6.1 User Interface Implementation

The USRAnalyzer prototype interfaces with consumers via interactive webpages, which provide users with user-friendly Graphical User Interface (GUI). Two examples of the user interface are given by Figure 6 and Figure 7 (see section 4.2), which present the interaction between the consumer and the USRAnalyzer. Figure 6 exemplifies the Home window of the user interface, and Figure 7 shows the Feedback window. The Home window provides the first interaction when a user runs the USRAnalyzer. The Feedback window displays additional interaction where the USRAnalyzer provides more flexibility to satisfy the need of the consumer. The Home and Feedback windows indicate that our USRAnalyzer prototype fulfills the research goal that the USRAnalyzer is an interactive web personalization artifact.

Additionally, the user interface interacts with the back end of the USRAnalyzer prototype. The interaction between the GUI and the back end is supported by seven web 2.0 technology frameworks: PHP, JavaScript, XHTML, XML, CSS, C++, and MySQL. PHP is a popular language for building server-side web application with the capability of dynamic web services. PHP enables the USRAnalyzer prototype to serve consumer need via the web server. XHTML and CSS have become the universal languages working hand-in-hand, which enables consumers to use the USRAnalyzer via their own web browsers (e.g. Internet Explorer, Mozilla FireFox, and Google Chrome). C++ is an advanced Microsoft .NET framework that can work with the other six frameworks seamlessly to offer personalized helpful reviews. MySQL is a popular language for creating and managing relational databases. The USRAnalyzer stores its structure data (data placed in tables) in MySQL database. XML is a framework for managing unstructured data (e.g. texts). XML is the most-used language for data exchange via the Internet. The USRAnalyzer manages its unstructured data in XML database.

Build on the seven advanced web 2.0 frameworks, the USRAnalyzer prototype can be implemented on any web server providing online reviews to personalize helpful reviews satisfying consumers' needs. Thus, the prototype indicates wide application of the USRAnalyzer.

6.2 Back End Implementation

We implement the back end by integrating C++ and R frameworks. They provide powerful functionality to the USRAnalyzer that can thus be implemented as web-based artifact with effective and efficient web processing capability. Powered by C++ and R, the USRAnalyzer

prototype can execute NLP tasks and machine learning tasks adequately to support web-based consumer decision support.

Moreover, the back end of the prototype can interact with both external lexicon repositories (e.g. WordNet or thesaurus.com) and internal lexicon repositories (e.g. the internal lexicon base that we built). From the trails-and-errors design process we went through, we have learned that an internal lexicon-base can help reduce Internet related issues that have undue influence on the performance of the USRAnalyzer. Internet issues include Internet connectivity issue, traffic jam, and security issue.

For the reasons described above, the USRAnalyzer prototype utilizes the internal lexicon base we built and used in our lab. For experimental evaluations, our internal lexicon base contains 132,761 relevant terms selected from thesaurus.com and WordNet. For the same reason, we downloaded 37,540 reviews on 8 services from tripadvisor.com and 100,460 reviews on 16 products from amazon.com. The selected reviews populate our internal review repository used in the experimental evaluations. When we downloaded the reviews, we removed the information associated with the reviews (e.g. author's names were removed). The associated information is not important for the USRAnalyzer.

Finally, the back end of the prototype implements our proposed DPSO-KM and RURanking algorithms as presented by the pseudo codes (refer to chapter 5, Figure 7 and Figure 8). We implement these two algorithms in R and C++. Particularly, all quantitative calculations

are implemented in R while the functions such as database connectivity and process control are implemented mainly in C++.

CHAPTER 7 EXPERIMENT EVALUATIONS

According to the Design Evaluation guideline proposed by Hevner et al. (2004), rigorous evaluation of design science (DS) artifact is critical for its utility. The evaluation requires appropriate metric and method.

In this chapter, we present the experimental evaluation of our research solution using the USRAnalyzer prototype (refer to Chapter 6 for its implementation). Specifically, we discuss the two experiments, which execute the evaluations of USRAnalyzer's effectiveness and efficiency. Using two large sets of real-world online reviews, the experimental evaluations contribute to literature what we believe, the first comprehensive evaluation of utility-sensitive review analysis (USRA) and web personalization (WP) artifacts.

The structure of this chapter is as following. In section 7.1, we present the overview of the two experiments. In section 7.2, we discuss the evaluation measures used in the two experiments. Then, we describe the first experiment, Experiment I, in section 7.3 and the second experiment, Experiment II, in section 7.4. In section 7.5, we discuss the implications of the two experiments.

7.1 Overview of Experiments

Two experiments were conducted to evaluate the performance of the USRAnalyzer using its prototype and real-world online reviews. Both experiments were designed to evaluate the effectiveness and efficiency of the proposed solution.

The first experiment used service reviews while the second experiment used product reviews. Thus, the evaluation assessed the utility of the USRAnalyzer to both service and product industries. We used simulated consumer inputs that were amounted to real-world consumer inputs (Geiger & Schader, 2014). Each simulated consumer input was corresponded to a gold relevant set (GRS), which is a relevant set RS amounted to the set of relevant reviews selected by real-world consumers. Web personalization research commonly uses a gold relevant set (GRS) as the standard to evaluate the accuracy of the relevant set selected by a web personalization artifact (Torkestani et al., 2012). In the absence of GRS, we asked 17 industrial experts to manually construct it for our experiments. Hereinafter, we refer to a simulated consumer input as c_h . We denote the GRS corresponding to c_h as GRS_h . It is equivalent to the high desirable relevant set that the real-world consumer will select from the review pool. Also, the USRAnalyzer prototype outputs a desirable relevant set. We denote it as the system desirable set (SDS_h). In the two experiments, the performance of the USRAnalyzer prototype was assessed by comparing SDS_h to GRS_h .

It is common for web personalization research to simulate consumer inputs and gold relevant sets for experimental evaluations (Geiger & Schader, 2014; Torkestani et al., 2012). Such approach is considered to be the best approach for producing reliable and generalizable evaluation results in web personalization research. Thus, we adopted such approach in the two experiments. Moreover, it is prohibitively expensive for us to collect a high volume of real-world consumer inputs. And, it is almost impossible for an actual consumer to manually identify the GRS from over thousand online reviews. Thus, to obtain adequate consumer inputs c_h and

reliable GRS_h , we have to construct them (Geiger & Schader, 2014; Ghorab et al., 2013; Shani & Gunawardana, 2011). The c_h and GRS_h were constructed by industry experts, and thus, helped improve the validation, reliability, and generalizability of our experimental evaluations.

7.2 Evaluation Metrics

This section discusses the evaluation metrics used in the two experiments, Experiment I (section 7.3) and Experiment II (section 7.4). According to the literature, web personalization research should evaluate two aspects related to the performance of proposed design science artifacts, effectiveness and efficiency (Lee & Kozar, 2012; Palmer, 2002). Effectiveness requires the USRAnalyzer to accurately identify the SDS_h . Efficiency requires the USRAnalyzer utilizes minimum resources to generate outputs. Prior web personalization research has mostly focused on effectiveness evaluation (Ghorab et al., 2013). In contrast, we evaluated both effectiveness and efficiency, using popular evaluation metrics in web personalization research.

The most-used metric for effectiveness evaluation is precision (p). Other frequently-used metrics include recall (r) and F-Measure (f) (Baeza-Yates & Ribeiro-Neto, 2011; Geiger & Schader, 2014). Precision (p) is the percentage of the retrieved reviews that are relevant to consumer need. Recall (r) is the percentage of the relevant reviews retrieved by the system. F-measure (f) is expressed as $f = 2pr/(p + r)$. We used two metrics p and f for three main reasons. First, precision is most used in web personalization research. Second, there are often trade-offs between precision and recall (Baeza-Yates & Ribeiro-Neto, 2011; Lee et al., 2012). Therefore, using precision only in experimental evaluations may be inadequate. To improve the validity of

our experimental evaluations, we used metric f since it could balance precision and recall. Using f enabled us to assess the ability of the USRAnalyzer to achieve optimal accuracy in terms of adequate precision and recall (Baeza-Yates & Ribeiro-Neto, 2011).

To evaluate the efficiency of the USRAnalyzer, we used the search time (t) metric, which is the time elapsed between submission of consumer request and output from the system. We do not count the time for consumer's activity (e.g., user inputting and user selecting). The t metric is equivalent to the metric of speed of data display that is the efficiency measure proposed by Lee et al. (2012). They also proposed other efficiency metrics such as navigation speed and page-loading speed. We did not use those metrics because they are more relevant to a complete web system. The USRAnalyzer serves as an analytical component of an entire website. The USRAnalyzer does not have control page-load and navigation speeds. Hence, the search time (t) is sufficient for the efficiency evaluation of the USRAnalyzer.

7.3 Experiment I

Experiment I was focused on USRAnalyzer's performance in handling service reviews. We used the online reviews of Rosen Inn International (RII) hotel, which were downloaded from tripadvisor.com, the most popular website for tourist services. In this section, we present the data collection used in Experiment I (section 7.3.1), the design of Experiment I (section 7.3.2), and the outcomes of Experiment I (section 7.3.3).

7.3.1 Experiment I Data Collection

Experiment I utilized the online reviews of Rosen Inn International (RII) hotel located in Orlando, Florida, USA. The review pool contained 3321 reviews of which the experiment I utilized 1227. We used a subset of the review pool for two reasons. First, the USRAnalyzer prototype has not implemented the Language Translation component of the proposed USRAnalyzer architecture (refer to section 4.3.1). Thus, the two experiments (Experiment I and Experiment II) were conducted by using online reviews written in English. The 3321 reviews in the RII review pool were not all written in English. Among them, only 1395 reviews are written in English. The review pool used in Experiment I came from the 1395 English reviews. Second, we avoided reviews having spelling errors because they could have undue influence on the evaluation results. Among the 1395 reviews, about 12% had spelling errors detected by the spelling checks. We rejected those erroneous reviews and obtained 1227 RII reviews that were written in English and adequate spelling.

As an additional note, we extracted only the content of the reviews and did not use reviewers' names and star rates accompanying with the reviews. Since the objective of the USRAnalyzer is to find helpful review content satisfying consumer needs, reviewers' names and star rates are irrelevant to the two experimental evaluations. Moreover, most review websites present the star rates and the summarized rates prominently. For example, tripadvisor.com provides rate summarizations for RII (Figure 17). A consumer can obtain them at a glance.

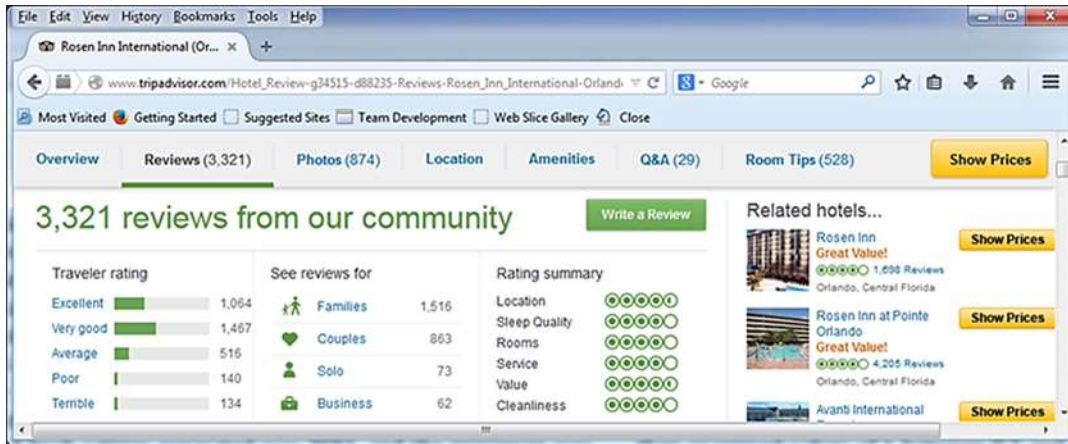


Figure 17 An Example of Star Rating at tripadvisor.com

7.3.2 Experiment I Design

Experiment I evaluated the performance of the USRAnalyzer when handling service reviews. In this experiment design, we included the expert-constructed consumer input c_h and the expert-constructed gold relevant set GRS_h . Combining with the experiment design, we conducted the tuning experiments to configure the experiment parameter.

Experiment I Consumer Input c_h : We worked together with seven marketing experts in tourist industry. They helped us simulate the consumer inputs by using Hotel Customer Experience Benchmarks (HCEB) of American Customer Satisfaction Index (ACSI). HCEB is not only an authoritative source for traveler needs, but also “the only uniform, cross-industry/government measure of customer satisfaction (Customer Satisfaction Study 2006)”. ACSI has established the HCEB via a series of surveys on hotel customer satisfaction. The surveys were conducted in 1994 through 2014. By analyzing the survey data, the HCEB has published the 10 hotel-consumer needs (HCNs) (Figure 18). The marketing experts used 6 HCNs

to construct the consumer inputs for Experiment I. They did not use in-room entertainment options, amenities, loyalty programs, and website primarily for two reasons. First, each of the first three HCNs are very ambiguous, so the experts' interpretations were open-ended. Second, the experts regarded website HCN as less relevant since a traveler can directly evaluate the hotel's website without using online reviews.

1. Ease of reservations
2. Check-in process
3. Room cleanliness and comfort
4. Food services
5. In-room entertainment options
6. Amenities
7. In-room Internet service
8. Staff courtesy
9. Loyalty programs
10. Website

Figure 18. HCEB's consumer needs

From the remaining 6 HCNs, the marketing experts constructed totally 62 distinct consumer inputs. A consumer input c_h involved one or more HCNs (from 4 to 10 in Figure 10). Figure 19 shows examples of eight consumer inputs. Input 1 through 6 each involves only one HCN. Input 7 involves two HCNs. Input 8 involves three HCNs. Also, the experts grouped the 62 consumer inputs into five input types by the number (k) of HCNs in an input. They denoted an input type as k -HCN ($k = 1, 2, 3, 4, 5$). To illustrate, each of c_1 through c_6 in Figure 19 belongs to 1-HCN while c_7 belongs to 2-HCN, and c_8 belongs to 3-HCN.

1. How easy is the reservation?
2. Is the check-in quick?
3. Is the room clean?
4. How comfortable is the bed?
5. Is there good food service?
6. The Internet connection should be good.
7. Polite staffs, comfortable room?
8. Easy reservation, room clearness, good food

Figure 19. Examples of Consumer Inputs

Experiment I Gold Relevant Set (GDS): For each c_h ($h = 1, 2, \dots, 62$) constructed by the marketing experts, we needed a gold relevant set GRS_h that amounts to the set of highly relevant reviews judged by the real-world traveler. We evaluated the accuracy of the system desirable set (SDS_h) output by the USRAnalyzer against the GRS_h . Since the GRS_h did not exist, the marketing experts helped us construct it manually. Although the construction process was manual, it was systematic and iterative. Also, the expert-constructed GRS_h resulted from a series of cross-checking. Therefore, the construction process was rigorous and supported the validation of the GRS_h as described below.

Before the GRS_h construction started, the experts randomly divided the 62 consumer inputs into seven input groups. Each of them contained 8 or 9 inputs. Then, they randomly distributed the input groups among them. Each expert executed a standard procedure to construct 8 or 9 proposals each for a GRS_h . The overall process involved two stages, construction and reconciliation.

The construction stage involved three iterative cycles. In the first cycle, each expert constructed the GRS_h proposals originally assigned to her/him. When everybody finished the assigned proposals, the seven input groups were randomly redistributed. This time, each expert received a different input group. Then, the second iterative cycle started during which each expert repeated the same procedure as the first cycle. When everybody finished the allocated proposals, the seven input groups are randomly redistributed a third time. And, each expert received a new input group different from the previous assigned groups, and repeated the same procedure as the previous cycles. Through three iterative cycles, the construction stage output three GRS_h proposals for each consumer input c_h .

In the reconciliation stage, the seven experts worked in two groups of 3 and 4. Each expert group compared the three GRS_h proposals for c_h . If they generally agreed, any of them became the output GRS_h . If there was a disagreement between any two proposals, the experts reconciled to address the inconsistency. The goal of the reconciliation was to ensure that the derived GRS_h was as close as possible to the desirable set selected by the real-world traveler. The experts developed 62 GDS_h after rigorous construction and reconciliation.

Experiment I Tuning Experiments: Parameter tuning is important for reliable evaluation of web personalization artifacts (Lee et al., 2012; Sanjay et al., 2013; Shi et al., 2014; Torkestani et al., 2012). For the USRAnalyzer, we tuned parameter δ , the cutoff point of the similarity between expanded consumer inputs and each relevant review. Parameter δ can significantly affect F-Measure (f) and Search Time (t). A high δ improves precision (p) but reduces recall (r). In general, a high δ causes the USRAnalyzer to reduce the size of relevant set

and reject more reviews with low relevance levels. The result can improve precision and efficiency, but can decrease recall. The rate of decreasing in recall is often higher than the rate of increasing in precision (Billerbeck, 2005; Yin et al., 2009). The trade-off between p and r can bias the f value. Therefore, we conducted careful tuning experiments before the evaluation experiment.

We started the tuning experiments by randomly sampling 250 reviews and by randomly sampling 15 consumer inputs. Two consumer inputs were of 1-HCN and 5-HCN types respectively, four were 2-HCN and 5-HCN, and six of 6-HCN. Then, we used the corresponding GRS_h to tune δ . For each δ value, we performed 15 runs. In each run, we used a c_h . Also, we averaged the performances of p , f , and t respectively over the 15 runs. The tuning experiments indicated that δ increased p steadily when δ was between 0.50 and 0.86. In that range, f decreased with relatively large margin, and t decreased marginally. When $\delta > 0.86$, f deteriorated noticeably. We set $\delta = 0.86$ for two reasons. First, p should be prioritized over f since web personalization systems stress precision (Lee et al., 2012; Yin et al., 2009). Second, high accuracy is not a difficult issue with low relevance threshold but a very difficult issue with high relevance level. In Table 4, we present the average results of the tuning experiments.

Table 4. Experiment I Tuning Experiments

δ	p	f	t (s)
0.50	0.871	0.858	0.062
0.55	0.872	0.851	0.061
0.62	0.875	0.842	0.061
0.68	0.879	0.833	0.058
0.77	0.882	0.825	0.057
0.86	0.886	0.651	0.055

7.3.3 Experiment I Outcomes

For each 62 consumer inputs, we ran the USRAnalyzer three times. Each time, we recorded the performance metrics p , f , and t . A total of 186 runs were performed. In each run, we randomly selected a consumer input c_h . After three runs using c_h , we averaged the three values of each performance metrics: p , f , and t . Thus, each metric had 62 averages. We grouped them by input type, i.e. k -HCN ($k = 1, 2, 3, 4, 5$). For each group, we computed the averages of p , f , and t respectively. The p , f , and t represented the performance of the USRAnalyzer at the complexity level of the input type, which is typically defined as the number of words in a consumer input. In general, a consumer input with one to four words has a low complexity. A consumer input with four to ten words has a medium complexity. And, a consumer input with ten to twenty words has high complexity (Billerbeck, 2005; Hauff et al., 2008). For our experiments, this implies that input type 1-HCN and 2-HCN are low complexity. 3-HCC is medium complexity, and 4-HCN and 5-HCN types are high complexity. Thus, our experiments covered the full range of input complexity for generalizable evaluation conclusions.

In Table 5, we present the performance outcomes from Experiment I. The performances on p and f increased steadily across the five input types. The stable increase implies the reliability of the USRAnalyzer across complexity levels of input types. Especially, the increase of p and f at high complexity of input type demonstrated the efficacy of the USRAnalyzer in real-world applications since actual consumer inputs are typically ten to twenty words (Billerbeck, 2005). As expected, the f values were lower than corresponding p values because we prioritized the performance on p . In general, the experiment results supported the effectiveness

of the USRAnalyzer in terms of high accuracy. It achieved high p and f across different input complexity-levels. Popular web personalization systems such as PNB and PEBL achieve p levels between 0.50 and 0.60 and f levels between 0.30 and 0.55 (Geiger & Schader, 2014; Lee et al., 2012). Moreover, the search time t generally increased along the increase in the complexity of the consumer inputs. That was expected since complex inputs require more computation. Web personalization researchers deem web personalization system as efficient when search time is less than 1 second on average (Teevan et al., 2013). Overall, the t values indicate that the USRAnalyzer is very efficient.

Table 5. Experiment I Outcomes

	2-W (1-HCN)	4-W (2-HCN)	6-W (3-HCN)	8-W (4-HCN)	10-W (5-HCN)	12-W (5-HCN)
p	0.819	0.827	0.828	0.836	0.839	0.841
f	0.752	0.754	0.755	0.757	0.761	0.762
t (s)	0.140	0.180	0.220	0.310	0.403	0.495

Note: 2-W for 2-word input, 4-W for 4-word input, 6-W for 6-word input, etc.

7.4 Experiment II

Experiment II was focused on USRAnalyzer’s performance in handling product reviews. We used the online reviews of Epson XP-310 Wireless Color Photo Printer, which were downloaded from amazon.com. In this section, we present the data collection used in Experiment II (section 7.4.1), the design of Experiment II (section 7.4.2), and the outcomes of Experiment II (section 7.4.3).

7.4.1 Experiment II Data Collection

We selected the product, Epson XP-310 Wireless Color Photo Printer (Epson XP-310). We downloaded its 1389 related reviews from amazon.com after rejecting the reviews with spelling errors. The reviews covered a broad range of consumer needs. Less than 9% of the 1389 reviews were very similar. The diversified consumer needs enforced the generalizability of Experiment II evaluation.

7.4.2 Experiment II Design

Experiment II Consumer Inputs: We obtained the help of ten customer service managers of Printer products. The managers simulated the consumer inputs using the professional printer reviews on pcworld.com and pcmag.com, which are authoritative online magazines for electronic industry. The professional printer reviews covered the broadest consumer concerns about printers. For consistency, each manager evaluated the same set of professional printer reviews and simulated 20 consumer inputs. Then, three of the ten managers worked together. They combined the 200 (10 x 20) consumer inputs and eliminated the redundant ones. Then, they identified totally 50 unique consumer inputs. Only 25% of the 200 consumer inputs were unique, which reflected high agreement among the ten managers. Then, the customer service managers validated the consumer inputs by using the survey data from the online survey of 70 printer consumers.

Experiment II Gold Relevant Sets: For each consumer input c_h out of the 50, we needed a gold relevant set (GRS_h , $h = 1, 2 \dots 50$) amounted to the real-world consumer would

select. The ten managers helped us build the GRS_h . The construction process was manual, but systematic, iterative, and rigorous. Consequently, each resultant GRS_h was reliable and valid. The construction process went through 3 stages as described next.

Stage I, constructing GRS_h proposals. The GRS_h construction started with the random division of the 1389 reviews into ten groups. Nine of these groups each contained 139 reviews while one group contained 138 reviews. Then, the construction entered three iterations. In the first iteration, the managers randomly divided the ten groups of the reviews among them. Then, each manager constructed 50 GRS proposals using the assigned reviews. In the second iteration, the 10 review groups are randomly redistributed. Each manager received a review group different from the one in the first iteration and repeated the tasks as in the first iteration. The third iteration is the same as the second. The difference was that each manager worked on a review group different from the previous two assignments.

Stage 2, proposing local GRS_h . After the three iterations, there were three GRS_h proposals for each c_h and each 10 review groups. In order to build one GRS_h from the three proposals, the 10 managers worked in pairs. They compared the three proposals in order to reconcile the difference. After reconciliation, they obtained one local GRS_h proposal for each c_h and each 10 review groups.

Stage 3, constructing output GRS_h . For each c_h , the managers combined the 10 local GRS_h proposals, and ended with 50 GRS_h ($n = 1, 2, \dots, 50$) respectively corresponding to the 50 c_h .

Experiment II Parameter Tuning: As in Experiment-I, we carried out a series of tuning experiments to tune parameter δ . We randomly sampled 260 reviews from the 1389 and 5 consumer inputs from the 50. We use the corresponding GRS_h . For each δ value, we ran the USRAnalyzer prototype five times. Each time, we used a different I_h . Also, we averaged the performances of p , f , and t respectively over the 5 runs. The tuning experiments indicated that increasing δ increased p steadily in the δ range from 0.50 to 0.93. In the same range, f decreased with a large margin, and t decreased marginally. When $\delta > 0.93$, f deteriorated noticeably. As in Experiment I, we prioritized the performance of p and t . We chose the highest threshold $\delta = 0.93$. Table 6 shows the average results.

Table 6. Experiment II Tuning Experiments

δ	p	f	t (s)
0.50	0.912	0.985	0.043
0.61	0.916	0.886	0.041
0.72	0.918	0.862	0.040
0.80	0.922	0.846	0.035
0.87	0.923	0.820	0.035
0.93	0.924	0.693	0.026

7.4.3 Experiment II Outcomes

For each c_h , we ran the USRAnalyzer prototype five times. Each time, we recorded the performances of p , f , and t . Total 250 runs were performed. We averaged the five performances for each c_h and each metric. Then, we averaged the 50 average performances for each performance metric, and then, aggregated p , f , and t performances respectively over c_h . Additionally, the 50 consumer inputs were distributed over three complexity levels: low (one- to four-word input), medium (four- to ten-word), and high (ten- to twenty-word input). The

complexity distribution of the 50 consumer inputs is 16 at low level, 20 at medium level, and 14 at high level. Thus, like Experiment I, Experiment II covered the full range of input complexity to draw generalizable evaluation conclusions.

In Table 7, we display the resultant performances from Experiment II. When the complexity of the consumer inputs increased, p and f generally increased. But, for 16- and 18-words inputs, p started to drop. The drop might be due to the increased complexity of the consumer inputs, which caused ambiguity among the noun-phrases. Such ambiguity increased the difficulty for the USRAnalyzer prototype to evaluate the relevance of the online reviews. Probably for the similar reason, f dropped when the consumer input contained 12, 16, and 18 words. Also, f dropped earlier than p . The reason might be because the decrease in recall augmented the decrease in f .

Nevertheless, Experiment II outcomes clearly show that the USRAnalyzer prototype achieved high p and f across different input complexity-levels. The slight drops occurred for complex consumer inputs did not affect the effectiveness and efficiency of the USRAnalyzer. Thus, Experiment II also supported the high accuracy of the USRAnalyzer. Finally, the search time t increased when input complexity increased. In Table 7, when consumer input contained 12 or more than 12 words, t jumped sharply. Such jump can be investigated carefully in future research that may advance our understanding of the USRAnalyzer (refer to chapter 8 of this paper). Thus, Experiment II provided review helpfulness research an interesting problem. However, according to the common understanding of efficient web personalization artifacts, the t

in relation to complexity of consumer inputs in Experiment II supported the efficiency of the USRAnalyzer.

Table 7. Experiment II Outcomes

	1-W	2-W	4-W	6-W	8-W	12-W	16-W	18-W
<i>p</i>	0.870	0.879	0.887	0.895	0.930	0.937	0.899	0.897
<i>f</i>	0.861	0.864	0.868	0.872	0.875	0.874	0.871	0.869
<i>t</i> (s)	0.09	0.10	0.13	0.14	0.17	0.32	0.43	0.48

Note: 1-W for 1-word input, 2-W for 2-word input, 4-W for 4-word input, etc.

7.5 Implications of Experiments

In this section, we summarize five implications from the two experiments presented in section 7.3 and section 7.4. First, the two experiments consistently show that the proposed USRAnalyzer can adequately address the research problem raised in section 1.3 of this dissertation: What is an effective and efficient interactive web personalization (IWP) artifact that can provide personalized utility-sensitive review analysis (PUSRA) meeting the changing needs of individual consumers? Particularly, the data analysis presented in Table 5 and Table 7 indicates that the proposed USRAnalyzer is an effective and efficient artifact for personalized utility-sensitive review analysis (PUSRA). The prototype demonstrated superior performance comparing to prior web personalization artifact in terms of efficiency and effectiveness (see section 7.3.3 and section 7.4.3).

Second, the two experiments exemplify one better approach to acquire reliable consumer inputs and gold relevant sets for experimental evaluations of utility-sensitive review analysis (USRA) and web personalization (WP) artifacts. As a relatively new research field, USRA and

WP research commonly relies on manually-constructed consumer inputs and gold relevant sets for experimental evaluations. The manual constructions enable the researchers to carry out reliable and generalizable evaluations (Geiger & Schader, 2014; Shani & Gunawardana, 2011). For instance, Liu et al. (2007) manually built a gold set of helpful reviews to investigate the problems of helpful votes on the review websites. Sugiyama et al. (2004) utilized 20 experienced web users to construct the gold relevant sets in order to compare three web personalization systems. However, prior research mostly used researcher-constructed user inputs and gold relevant sets rather than expert-constructed ones. To the best of our knowledge, our experiments are the first review helpfulness research that utilizes expert-construction approach. This approach adds more practicality and validity to the gold relevant sets. Industry experts have intimate knowledge and experience on industry trends and customer use of personalization systems. Researchers often lack such knowledge and experience. In the Experiment I, the marketing experts helped us appropriately utilize industry authoritative survey data. In the Experiment II, the customer service managers validated the consumer inputs by using survey data. Thus, expert-construction approach substantially improved the validation of USRAnalyzer's practical utility, as well as the reliability and generalizability of the two experimental evaluations.

Third, efficient evaluation of WP artifacts is fundamental to their practical usability (Yin et al., 2009). A few WP researches conducted efficiency evaluation (e.g. He & Ounis, 2007). But, prior USRA research has commonly missed efficiency evaluation. We stressed both effectiveness and efficiency in our two experiments and demonstrated the effectiveness and the efficiency of the USRAnalyzer. Specifically, the Experiment II indicated that artifact efficiency

should be a focus of USRA and WP research. There was a marked jump in search time when consumer inputs became complex (refer to section 7.4.3). This suggests that future research should place efficiency as an important issue. Our research presents new opportunities for USRA and WP researchers to address efficiency bottlenecks so that research contributions become more relevant to businesses, web technology professionals, and consumer communities.

Fourth, the Experiment I and II demonstrated that the USRAnalyzer improved retrieve accuracy 15% to 20% compared to prior WP research. So, we consider the USRAnalyzer effectiveness (i.e. $0.50 < p < 0.60$ and $0.30 < f < 0.55$) favorable to WP effectiveness (Geiger & Schader, 2014; Lee et al., 2012). Additionally, we regard the USRAnalyzer is reliable. In the two experiments, we used the highest relevance thresholds ($\delta = 0.86$ and 0.93). High effectiveness is a more difficult goal under a high relevance threshold compared to low relevance threshold. Consequently, the two experiments indicated that our proposed algorithms of DPSO-KM and RURanking together contributed to USRAnalyzer's effectiveness. This provides the well-known tenet, 'effective WP artifacts require the combined effectiveness of all the algorithms'.

Fifth, the outcomes of the two experiments indicated that the USRAnalyzer could more effectively handle product reviews than service reviews. The p and f outcomes of the Experiment II were respectively higher than those of the Experiment I. Also, the relevance threshold of the Experiment II was higher than that of the Experiment I. The performance difference could be from many reasons. For example, the reviews of RII hotel might be less distinguishable than those of Epson XP-310 printer. Or, the overall writing quality of Epson XP-310 reviews was higher than that of RII reviews. For our current research, the possible reasons for the

performance difference are out of the scope. However, those reasons are certainly interesting issues for future research to explore. An USRAnalyzer that is equally effective for product and service reviews can be easier to implement and maintain.

CHAPTER 8 FUTURE RESEARCHES

The two experimental evaluations presented in Chapter 7 suggest that our proposed solution has achieved the research objectives. Following from this initial success, our future research will address five challenging problems discussed in this chapter.

The first problem is implementing the Language Translator component presented in the architecture of the USRAnalyzer (section 4.3). The multi-lingual capability is important to the broad application of the USRAnalyzer, which should be extended to being able to work with any review website on the Internet regardless the natural language used by the reviews on the website. In today's globalized interconnectivity via the Internet, many products and services have worldwide consumers. At a single review website, the products and services may receive online reviews in multiple natural languages. For example, at tripadvisor.com, Rosen International Inn has accumulated online reviews in English, German, French, Korean, Chinese, etc. The quantity of the reviews in each language is large, i.e. hundreds at minimum. That also suggests that worldwide consumers may use online reviews at a single review website. Therefore, a multi-lingual USRAnalyzer can greatly benefit all consumers worldwide to get valuable helpful reviews. Thus, multi-lingual reviews can be viewed to better support consumers' purchase decisions. For such reason, we attempt to address in the immediate future, this consumer need by empowering the USRAnalyzer with the most advanced language translation technology.

The second problem is evaluating practical effectiveness of the proposed USRAnalyzer in its real-world usages. Although the two experiments presented in chapter 7 strongly supported the practical utility of the USRAnalyzer, its practical utility needs more scrutiny in daily usages on review websites. The real-world applications may expose important improvement needed to be addressed in the future. We plan to work with business industry to utilize the USRAnalyzer on several review websites. We will collect and analyze the usage data for further evaluations of effectiveness and efficiency. And, we will study broadened user base of the USRAnalyzer to validate its capability in the real-world environment. Such study will provide valuable knowledge not only for us to improve the current solution but also for other researchers to create better artifacts with similar functionality as that of the USRAnalyzer.

The third problem is to evaluate the scalability of the USRAnalyzer and the review websites using the USRAnalyzer. The real-world applications discussed in the second future research problem will enable us to evaluate the scalability of the USRAnalyzer-enabled review websites. Since the bodies of online reviews and their users are expected to growth exponentially, scalability is critical to broad application of the USRAnalyzer-enabled review websites because they need to handle, without bottleneck, thousands or millions user activities at the same time. Similar to addressing the second problem, we will work with business industry to implement multi-lingual USRAnalyzer on review websites. We will collect and analyze the data to evaluate the scalability of the USRAnalyzer and the review websites. According to the evaluation, we will improve the USRAnalyzer if scalability issue is discovered.

The fourth problem is enabling the USRAnalyzer to work with multiple review websites and providing personalized utility-sensitive review analysis (PUSRA) by identifying helpful reviews of a product or a service from multiple review web websites. Such capability is very important to consumers and businesses. They are facing the problem of exploding growth of review websites and rely on automatic solutions to find useful online reviews aggregated from multiple review websites. The helpful reviews pooled from different websites provide additional value and rich opinions that the helpful reviews from a single website may not offer (Herlocker et al., 2004). In order to work with multiple review websites, the USRAnalyzer needs to incorporate crowd computing that is the computational capability enabling powerful search engine (e.g., Google and Yahoo) to perform multi-website search with effectiveness and efficiency. The need for a multi-website USRAnalyzer will bring a series of new challenges to research community, including how to identify relevant review websites, how to assess the relevancy of the online reviews across multi-review pools, and how to evaluate a future-USRAnalyzer empowered by cloud computing. It is certain that promising research solution can be created along with the growth of clouding computing technology.

The fifth problem is comparing the performance of the USRAnalyzer to other personalized utility-sensitive review analysis (PUSRA) approaches. Currently, such comparison is very difficult, if possible, because the existing PUSRA approach (e.g. Moghaddam et al., 2012) has not provided a design science (DS) artifact. The USRAnalyzer is, what we believe, the first DS artifact prototyped for PUSRA. However, along with the increasing research effort on PUSRA, we expect to see future contributions of DS artifacts for PUSRA. We can thus compare

the performance of the USRAanalyzer and that of the future PUSRA artifacts. The comparison can offer substantial learnings for advancing review helpfulness research (RHR) and web personalization research (WPR) (Ghorab et al., 2013). Also, the comparison can help improve web-based consumer support research where an interesting problem is: whether PUSRA artifacts can add more product sales comparing to other USRA approaches. The study of such problem can reveal in-depth understanding of the utility of PUSRA artifacts.

According to the discussion given above, the readers may see that our proposed USRAnalyzer initializes an exciting and broad research future for review helpfulness research, web personalization research, and web-based consumer support research. The future research directions presented in this chapter are far less than exhaustive. Our discussion of future research in this chapter covers only a small portion of the promising future research. And, the future review helpfulness research will most likely go beyond the limits.

CHAPTER 9 CONCLUSIONS

Online reviews have become one of the important resources for consumers' purchase decisions. Many consumers regard online reviews highly in their purchase decision making. For online reviews have great influence on consumers' purchase choices, businesses utilize online reviews to understand their customers. In utilizing online reviews, an important concern to consumers and businesses is the helpfulness, usefulness, or utility of an online review to improve consumer purchase decision. Such concern is referred to as review utility, which reflects the fact of the equivocal utility of online reviews in terms of helping consumers make better purchase decisions.

The equivocal utility stresses the fact that only helpful reviews can help improve purchase decisions. However, identifying them can be a very difficult problem to consumers when automatic approach is not available. This problem facing consumers calls for automatic approaches for utility-sensitive review analysis (USRA), which is the automatic process to assess the helpfulness of a review for improving consumer purchase decisions.

In this dissertation, we have demonstrated that the helpfulness of a review needs to be personalized. An online review must satisfy a specific need of a specific consumer in order to be useful to that consumer. Therefore, USRA approaches require a web-based and personalized solution, which in turn needs an interactive web personalization (IWP) artifact. Unfortunately, literature does not offer it, and personalized USRA (PUSRA) is not available in practice. The

existing USRA solutions are limited to the predictive models, which are largely inadequate for PUSRA. To fill the literature gap, our current research addresses this difficult problem by proposing and evaluating the USRAnalyzer, which is an interactive web personalization artifact.

The two experiments presented in chapter 7 demonstrate that the USRAnalyzer can satisfy personalized consumer need via interacting with consumers. According to the best knowledge we have, the USRAnalyzer is the first design science (DS) artifact grounded on IS design science principle to achieve our research objectives (1) contributing substantial new knowledge to DS research in general, to review helpfulness research (RHR) and web personalization research (WPR) in particular; (2) maximizing the value of online reviews to consumers and businesses. Our evaluations of the USRAnalyzer prototype supported the achievement of our research objective. Thus, the USRAnalyzer can be said to adequately address the two essential problems of review helpfulness research: (1) minimizing information overloading to consumers who turn to online reviews for useful opinions; (2) minimizing review utility misrepresentation of the online reviews. In chapter 1 and chapter 2, we demonstrated that suitable solution for those two problems is critical for the benefit of consumers who utilize online reviews to improve their purchase decisions. At the time, suitable solution for those two problems is also crucial to businesses who utilize online review to understand their customers.

In the previous chapters, we presented the USRAnalyzer in detail. We described its overview model and architecture in chapter 4 and two algorithms (DPSO-KM and RURAlgorithm) in chapter 5. We discussed the implementation of the USRAnalyzer prototype in chapter 6. In chapter 7, we evaluated the proposed solution experimentally and

comprehensively in its effectiveness and efficiency. The readers may recall that the Experiment I and II presented in chapter 7 were executed rigorously with the help of the industry experts. The two experiments consistently corroborated that the USRAnalyzer can achieve high effectiveness comparing to prior web personalization (WP) solution. And, the USRAnalyzer is highly efficient in terms of established WP efficiency.

Therefore, our current research has achieved its five objectives. The USRAnalyzer contributes to review helpfulness research (RHR) the first WP solution, which initializes a new research direction for RHR. Second, the USRAnalyzer presents a theoretical contribution to the knowledge body of RHR and WP research, as well as design science (DS) and information systems (IS) research. In chapter 3, we presented that a practical useful DS artifact contributes to DS and IS knowledge body (Hevner et al., 2004). Our evaluations strongly support the practical utility of the USRAnalyzer. Moreover, our proposed DPSO-KM and RURAlgorithm contribute the operationalized methods that IS researchers and professionals can adopt to significantly improve their works. Researchers can continue improving the two algorithms for more advance WP solutions. Website designers can utilize the two algorithms to fundamentally improve the designs of review websites. Such improvement can help bring about substantial growth of review websites and business revenues.

In summary, this research has made none-trivial contributions to solve the unsolved research problem, i.e. an effective and efficient personalized utility-sensitive review analysis (PUSRA) artifact. Our contributions will bring about prolific USRA and WP research and practice. In the future, we will address the five research problems discussed in Chapter 8. We

will first implement the Language Translator component of the USRAnalyzer and evaluate its utility with its full functionality. Second, we will implement the fully functional USRAnalyzer on real-world review websites and evaluate its effectiveness and efficiency in its real-world usages. Moreover, we will continue studying to address the other three future research problems. And, our research experience in the design of the USRAnalyzer will contribute very important knowledge to the success of our future research. We will continue making substantial contributions to USRA and WP research and practice as well as IS research and professional practice.

LIST OF REFERENCES

LIST OF REFERENCES

- Adomavicius G., Sankaranarayanan R., Sen S., and Tuzhilin A. (2005) Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Transactions on Information Systems* 23, 1 (January 2005), pp. 103-145.
- Agarwal R., V. Venkatesh, Assessing a firm's web presence: a heuristic evaluation procedure for the measurement of usability, *Information Systems Research* 132 (2002) 168–186.
- Ambati V., Uppuluri, R. (2006) Using monolingual clickthrough data to build cross-lingual search systems. *New Directions in Multilingual Information Access Workshop of SIGIR 2006*, ACM, Seattle (2006).
- American Customer Satisfaction Index (ACSI)
(<http://www.theacsi.org/industries/travel/hotel>)
- Amorim R. C., Mirkin, B (2012) Minkowski metric, feature weighting and anomalous cluster initializing in K-Means clustering. *Pattern Recognition*, 45 (3), pp. 1061–1075.
- Arguello J., Elsas J. L., Callan J., Carbonell J. G. (2008) Document Representation and Query Expansion Models for Blog Recommendation. *ICWSM*, 2008(0), 1.
- Atahan P, Sarkar S (2011) Accelerated learning of user profiles. *Management Science*, 57(2), pp. 215–239.
- Baeza-Yates R., Ribeiro-Neto B. (2011) *Modern Information Retrieval: The Concepts and Technology Behind Search*, 2nd edn. Addison-Wesley, Reading (2011).
- Baeza-Yates R., Ribeiro-Neto B. (1999). *Modern information retrieval* (Vol. 463). New York: ACM press.
- Bakos J.Y. (1997) Reducing buyer search costs: Implications for electronic marketplaces. *Management Science*, 43(12), pp. 1676-1692.

- Baltrunas L., Makcinskas T., Ricci F. (2010) Group recommendations with rank aggregation and collaborative filtering. In Proceedings of the 4th ACM Conference on Recommender Systems (RecSys'10). ACM, New York, NY, pp. 119–126.
- Basole R. C., Seuss C. D., Rouse W. B. (2013) IT innovation adoption by enterprises: Knowledge discovery through text analytics, *Decision Support Systems*, 54 (2013), pp. 1044–1054.
- Barros R.C. Ruiz D.D., Basgalupp M. (2011) Evolutionary model trees for handling continuous classes in machine learning, *Information Sciences*, vol. 181, pp. 954–971, 2011.
- Baskerville R.L., Kaul M., Storey V.C. (2011) Unpacking the Duality of Design Science. In Proceeding of Thirty Second International Conference on Information Systems, Shanghai 2011.
- Bast H., Majumdar D., Weber I. (2007) Efficient interactive query expansion with complete search. In: 16th ACM Conference on Information and Knowledge Management (CIKM 2007), pp. 857–860. ACM, Lisbon (2007).
- Beeferman D., & Berger, A. (2000) Agglomerative clustering of a search engine query log. In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 407-416). ACM.
- Berry M.A., Linoff G. (2010) *Data Mining Techniques, For Marketing, Sales, and Customer Relationship Management (5 Ed)*. Wiley Publishing, Inc. (2010).
- Bhargava H.K., Power D.J., Sun D. (2007) Progress in Web-based decision support technologies. *Decision Support Systems* 43 (2007), pp. 1083– 1095.
- Billerbeck B., Scholer F., Williams H. E., Zobel J. (2003). Query expansion using associated queries. In Proceedings of the twelfth international conference on Information and knowledge management (pp. 2-9).
- Billerbeck B., Zobel J. (2004) Techniques for efficient query expansion. In *String Processing and Information Retrieval*, Springer Berlin Heidelberg, pp. 30-42.

- Bina B., Schulte O., Crawford B., Qian Z., Xiong Y. (2013) Simple decision forests for multi-relational classification, *Decision Support Systems*, 54 (2013), pp. 1269–1279.
- Bo Geng B., Linjun Yang, Chao Xu, Xian-Sheng Hua (2012) Ranking Model Adaptation for Domain-Specific Search. *Knowledge and Data Engineering, IEEE Transactions*, 2012, pp. 745-758.
- Bodapati A.V. (2008) Recommendation systems with purchase data, *Journal of Marketing Research (JMR)* 45 (1) (2008) 77–93.
- Cagnina L., Errecalde M., Ingaramo D., Rosso P. (2014). An efficient Particle Swarm Optimization approach to cluster short texts. *Information Sciences*, 265, pp. 36-49.
- Cao G., Gao J. Nie J.-Y., Bai J. (2007) Extending query translation to cross-language query expansion with Markov chain models. In: 14th ACM International Conference on Information and Knowledge Management (CIKM 2007), pp. 351–360. ACM, Lisbon (2007).
- Cao Q., Duan W., Gan Q. (2011) Exploring determinants of voting for the “helpfulness” of online user reviews: a text mining approach. *Decision Support Systems* (50) (2011), pp. 511–521.
- Cenfetelli R, Benbasat I, Al-Natour S (2008) Addressing the what and how of online services: Comparing service content and service quality for e-business success. *Information Systems Research*, 19(2):161–181.
- Chang Y., Ounis I., Kim M. (2006). Query reformulation using automatically generated query concepts from a document space. *Information processing & management*, 42(2), pp. 453-468.
- Chen C.C., Tseng Y.-D. (2011) Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, (50) (2011) 755–768.
- Chen Y., Xie, J., Hall M. C. (2008) Online consumer review: word-of-mouth as a new element of marketing communication mix, *Management Science* (54:3) (2008) 477–491.

- Cheung C.M., Lee M.K. (2012) What drives consumers to spread electronic word of mouth in online consumer-opinion platforms. *Decision Support Systems* (53:1) (2012), pp. 218–225.
- Chevalier J.A., Mayzlin D. (2006) The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, August 2006, Vol. 43, No. 3, pp. 345-354.
- Chire (2010) *k*-means clustering result for the Iris flower data set at http://en.wikipedia.org/wiki/Iris_flower_data_set.
- Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to information retrieval*, Vol. 1, Cambridge university press, 2008.
- Chung W., Chen (2005) A visual knowledge map framework for the discovery of business intelligence on the Web. *Journal of Management Information Systems* 21, 4, pp. 57–84.
- Clemons E.K., G. Gao, L.M. Hitt (2006) When online reviews meet hyper differentiation: a study of craft beer industry, *Journal of Management Information* 23 (2) (2006), pp. 149–171.
- Clemons EK (2008) How information changes consumer behavior and how consumer behavior determines corporate strategy. *Journal of Management Information Systems*, 25(2), pp.13–40.
- Cobos C., Henry Muñoz-Collazos, Richar Urbano-Muñoz, Martha Mendoza, Elizabeth León, Enrique Herrera-Viedma (2014) Clustering of web search results based on the cuckoo search algorithm and Balanced Bayesian Information Criterion. *Information Sciences*, Volume 281, 10 October 2014, pp. 248-264.
- Compeau D., Marcolin B., Kelley H., Higgins C. (2012) Generalizability of Information Systems Research Using Student Subjects – A Reflection on Our Practices and Recommendations for Future Research. *Information Systems Research*, Vol. 23, No. 4, December 2012, pp. 1093-1109.
- Cui H., Wen J. R., Nie J. Y., Ma W. Y. (2003). Query expansion by mining user logs. *Knowledge and Data Engineering, IEEE Transactions on*, 15(4), pp. 829-839.

- Cyr D., M. Head, H. Larios, B. Pan, Exploring human images in website design: a multi-method approach, *MIS Quarterly* 33 (2009) 539–566.
- Danescu-Niculescu-Mizil C., Kossinets G., Kleinberg J., Lee L. (2009) How opinions are received by online communities: a case study on amazon.com helpfulness votes. In *Proceeding WWW '09 Proceedings of the 18th international conference on World wide web*, pp. 141-150.
- Dasgupta S., Freund, Y. (2009) Random Projection Trees for Vector Quantization. *Information Theory, IEEE Transactions on* 55: 3229–3242.
- David S., Pinch T.J. (2006) Six degrees of reputation: The use and abuse of online review and recommendation systems. *First Monday*, July 2006.
- DeBono, K. G., A. Leavitt, J. Backus. (2003) Product packaging and product evaluation: An individual difference approach. *Journal of Application Social Psychology*, 33(3), pp. 513–521.
- Degeratu A, Arvind R, Wu J (2000) Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of Research in Marketing*, 17(1), pp. 55-78.
- Dellarocas C. (2003) The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10), pp. 1401-1424.
- DeLone W.D., E.R. McLean, The DeLone and McLean model of information systems success: a ten-year update, *Journal of Management Information Systems* 19 (2003) 9–30.
- Dong J., Du H.S. Wang S., Chen K., Deng X. (2004) A framework of Web-based Decision Support Systems for portfolio selection with OLAP and PVM. *Decision Support Systems* 37 (2004), pp. 367– 376.
- Dou Z., Song R., Wen J.-R., Yuan X. (2009) Evaluating the effectiveness of personalized Web search. *IEEE Transaction of Knowl. Data Eng.* 21, 1178–1190 (2009).

- Duan W., Gu B., Whinston A.B. (2008) Do online reviews matter? — An empirical investigation of panel data, *Decision Support Systems*, 45 (4) (2008), pp. 1007–1016.
- Eroglu S. A., Machleit K. A., Davis L. M. (2003) Empirical testing of a model of online store atmospherics and shopper responses. *Psychology and Marketing*, 20, 2, 2003, pp. 139–150.
- Evrin V., McLeod D. (2014) Context-based information analysis for the Web environment. *Knowledge and Information Systems*, January 2014, Volume 38, Issue 1, pp 109-140.
- Fawcett Y., *ROC Graphs: Notes and Practical Considerations for Researchers*, Machine learning, 2004.
- Forgy E.W. (1965) Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, **21**, pp. 768–769.
- Freund Y., Schapire R.E. (1997) A decision-theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.
- Furnas G. W., Landauer T. K., Gomez L. M., Dumais S. T. (1987) The vocabulary problem in human-system communication. *Comm. ACM* 30, 11, 964–971.
- Gao W., Niu C., Nie J.-Y., Zhou D., Hu J., Wong K.-F., Hon H.-W.: Cross-lingual query suggestion using query logs of different languages. In: 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007), pp. 463–470. ACM, Amsterdam (2007).
- Geem Z., J. Kim J., Loganathan G.V. (2001) A new heuristic optimization algorithm: harmony search. *Simulation*, 76 (2001) 60–68.
- Geige D.r, M. Schader (2014) Personalized task recommendation in crowdsourcing information systems — Current state of the art, *Decision Support Systems* 65 (2014), pp. 3–16.

- Ghorab M.R., Zhou D., O'Connor A., Wade V. (2013) Personalised Information Retrieval: survey and classification. *User Model User-Adaptive Interface*, (2013) 23, pp. 381–443.
- Ghose A., Ipeirotis P. G. (2010) Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 2010.
- Ghose A., Ipeirotis P.G. (2007) Designing novel review ranking systems: Predicting usefulness and impact of reviews. in *Proceedings of the International Conference on Electronic Commerce (ICEC)*, 2007.
- Gilbert E., Karahalios K. (2010) Understanding deja reviewers. *CSCW2010*, ACM, Savannah, Georgia, USA, 2010, pp.225–228.
- Goes P.B. (2014) Editor's Comments, *Design Science Research in Top Information Systems Journals*. *MIS Quarterly*, Vol. 38 No. 1, March 2014, pp. iii-viii.
- Gordon M., M. Kochen (1989) Recall-Precision trade-off: A derivation. *Journal of the American Society for Information Science*, 40, 1989, pp. 145-151.
- Granados N, Gupta A, Kauffman R (2012) Online and offline demand and price elasticities: Evidence from the air travel industry. *Information Systems Research*, 23(1), pp. 164–181.
- Gu B., Jaehong Park, Prabhudev Konana (2008) Research Note—The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products. *Information Systems Research*, Volume 23 Issue 1, March 2012, pp. 182-196.
- Haddoud M., Abdeddaïm S. (2014) Accurate keyphrase extraction by discriminating overlapping phrases, *Journal of Information Science*, April 15, 2014, published online.
- Haubl G., Trifts V. (2000) Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids. *Marketing Science*. Vol. 19, No. 1, Winter 2000, pp. 4–21.

- Hahm G.J., Mun Yong Yi, Jae Hyun Lee, Hyo Won Suh (2014) A personalized query expansion approach for engineering document retrieval. *Advanced Engineering Informatics*, Available online 16 May 2014.
- Hamerly G., Elkan C. (2004) Learning the k in A> means. *Advances in neural information processing systems*, 16 (2004), pp. 281.
- Hamilton, L. (2012). *Statistics with Stata: version 12*. Cengage Learning.
- Hauff C., Hiemstra D., de Jong F. (2008, October). A survey of pre-retrieval query performance predictors. In *Proceedings of the 17th ACM conference on Information and knowledge management* (pp. 1419-1420). ACM.
- Herlocker J.L., J.A. Konstan, L.G. Terveen, J.T. Riedl (2004) Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems* 22 (2004), pp. 5–53, <http://dx.doi.org/10.1145/963770.963772>.
- Hevner A. R., March S.T., Park J., Ram S. (2004). *Design Science In Information Systems Research*. *MIS Quarterly* Vol. 28 No. 1, March 2004, pp. 75-105.
- Hevner A.R. (2007) A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, (19:2), pp. 87-92.
- Ho S.Y., Bodoff D. (2014) The Effects of Web Personalization on User Attitude and Behavior: An Integration of the Elaboration Likelihood Model and Consumer Search Theory. *MIS Quarterly* Vol. 38 No. 2, pp. 497-520/June 2014.
- Hoang L., Lee J.-T., Song Y.-I., Rim H.-C. (2008) A model for evaluating the quality of user-created documents, in: H. Li, T. Liu, W.-Y. Ma, T. Sakai, K.-F. Wong, G. Zhou (Eds.), *AIRS 2008, LNCS 4993*, Springer-Verlag, 2008, pp. 496–501.
- Hofmann T. (1999) Probabilistic Latent Semantic Indexing. In *Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval*, 1999.

- Hong W. Y., Thong J. Y. L., Tam K. Y. (2004) Does Animation Attract Online Users' Attention? The Effects of Flash on Information Search Performance and Perceptions. *Information Systems Research*, (15:1), 2004, pp. 60-86.
- Hu J., Deng W., Guo J. (2006) Improving retrieval performance by global analysis. In *Pattern Recognition, ICPR 2006, 18th International Conference on Information Retrieval*, Vol. 2, pp. 703-706.
- Huang C. K., Chien L. F., Oyang Y. J. (2003) Relevant term suggestion in interactive web search based on contextual information in query session logs. *Journal of the American Society for Information Science and Technology*, 54(7), pp. 638-649.
- Hurst M., Nigam K. (2004) Retrieving topical sentiments from online document collections. *Document Recognition and Retrieval XI*, 2004, pp. 27-34.
- Iivari J. (2003) Towards Information Systems as a Science of Meta-artifacts. *Communications of the AIS*, (12)37, Nov., pp. 568-581.
- Iivari J., Hirschheim R., Klein H. K. (1998) A Paradigmatic Analysis Contrasting Information Systems Development Approaches and Methodologies. *Information Systems Research*, (9)2, pp. 164- 193.
- Jain A. K. (2010) Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31(8), pp. 651-666.
- Jiang Z., Benbasat I. (2007) The effects of presentation formats and task complexity on online consumers' product understanding. *MIS Quarterly*, 31(3), pp. 475–500.
- Jiang, J., Gretzel, U., Law, R. (2010) Do negative experiences always lead to dissatisfaction?—testing attribution theory in the context of online travel reviews. *Information and Communication Technologies in Tourism*, 2010, pp. 297–308.
- Jiang Y., Shang J., Liu Y. (2010) Maximizing customer satisfaction through an online recommendation system: A novel associative classification model. *Decision Support Systems*, 48(3), 470-479.

- Johar M., Mookerjee V., Sarkar S. (2014) Selling vs. Profiling: Optimizing the Offer Set in Web-Based Personalization. *Information Systems Research* 25(2), pp. 285-306.
- Kacem A., Boughanem M., Faiz R. (2014) Time-Sensitive User Profile for Optimizing Search Personalization. *Computer Science*, Volume 8538, 2014, pp 111-121.
- Kamal N., Hurst M. (2004) Towards a robust metric of opinion. In *Proceedings of AAAI Spring Symp. on Exploring Attitude and Affect in Text*. 2004.
- Kardan A.A., Ebrahimi M. (2013) A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups. *Information Sciences*, Volume 219, 10 January 2013, pp. 93-110.
- Kim M., Lennon S. (2008) The effects of visual and verbal information on attitudes and purchase intentions in internet shopping. *Psychology & Marketing*, Volume 25, Issue 2, February 2008, pp. 146–178.
- Kim S.-M., Pantel P., Chklovski T., Pennacchiotti M. (2006) Automatically assessing review helpfulness. in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 423–430, Sydney, Australia, July 2006.
- Kirkpatrick S., C. D. Gelatt Jr C.D., Vecchi M.P. (1983) Optimization by simulated annealing. *Science*, 220 (1983), pp. 678-680.
- Kohli R., S. Devaraj M. A. (2004) Understanding determinants of online consumer satisfaction: A decision process perspective. *Journal of Management Information Systems* 21(1), pp. 115–136.
- Komiak S.Y.X., Benbasat I. (2006) The Effect of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. *MIS Quarterly* Vol. 30 No. 4 pp. 941-960/December 2006.
- Kopliku A., Pinel-Sauvagnat K., Boughanem M. (2014) Aggregated search: A new information retrieval paradigm. *ACM Computing Surveys*, 46(3), 41.

- Koren Y. (2010) Collaborative filtering with temporal dynamics. *Communications of the ACM* 53, 4 (April 2010), 89–97.
- Koren Y., Bell R., Volinsky C. (2009) Matrix factorization techniques for recommender systems. *Computer*, 42(8), 2009, pp. 30–37.
- Korfiatis N., García-Bariocanal E., Sánchez-Alonso S. (2012) Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, 11 (2012), pp. 205–217.
- Kraaij, W., Nie, J. Y., & Simard, M. (2003). Embedding web-based statistical translation models in cross-language information retrieval. *Computational Linguistics*, 29(3), 381-419.
- Kraft, R., & Zien, J. (2004) Mining anchor text for query refinement. In *Proceedings of the 13th international conference on World Wide Web* (pp. 666-674). ACM.
- Ku Y.-C. Wei C.-P., Hsiao H.-W. (2012) To whom should I listen? Finding reputable reviewers in opinion-sharing communities. *Decision Support Systems* (53:3) (2012), pp. 534–542.
- Kuechler W., Vaishnavi V. (2008) The Emergence of Design Science Research in Information Systems in North America. *Journal of Design Research*, (7:1), pp. 1-16.
- Kuechler W., Vaishnavi V. (2012) A Framework for Theory Development in Design Science Research: Multiple Perspectives. *Journal of the AIS* (13:6), pp. 395-423.
- Kumar H., Lee S., Kim H.-G. (2014) Exploiting social bookmarking services to build clustered user interest profile for personalized search. *Information Sciences*, Volume 281, 10 October 2014, pp. 399–417.
- Kuruzovich J., Siva Viswanathan, Ritu Agarwal, Sanjay Gosain, Scott Weitzman (2008) Marketplace or Marketplace? Online Information Search and Channel Outcomes in Auto Retailing. *Information Systems Research*. Volume 19 Issue 2, June 2008, pp. 182-201.
- Lappas T, Gunopulos D. (2010) Efficient confident search in large review corpora. in *Proceedings of ECML-PKDD 2010*.

- Lee Y. E., Benbasat, I. (2011) Effects of Attribute Conflicts on Consumers' Perceptions and Acceptance of Product Recommendation Agents: Extending the Effort–Accuracy Framework. *Information Systems Research*, (22:4), pp. 867-884.
- Lee Y., Kozar K.A. (2012) Understanding of website usability: Specifying and measuring constructs and their relationships, *Decision Support Systems* 52 (2012), pp. 450–463.
- Lee Y.-H., P.J.-H. Hu, T.-H. Cheng, Y.-F. Hsieh (2012) A cost-sensitive technique for positive-example learning supporting content-based product recommendations in B-to-C e-commerce, *Decision Support Systems* 53 (2012), pp. 245–256.
- Leveling J., Jones G.J.F. (2010) Classifying and filtering blind feedback terms to improve information retrieval effectiveness. In: *Adaptivity, Personalization and Fusion of Heterogeneous Information (RIAO 2010)*, pp. 156–163. Le Centre De Hautes Etudes Internationales D'Informatique Documentaire, Paris (2010).
- Li T., Kauffman R.J., Heck E., Vervest P., Dellaert B.G.C. (2014) Consumer Informedness and Firm Information Strategy. *Information Systems Research*, 25(2), pp.345-363.
- Li X., Hitt L.M. (2008) Self-selection and information role of online product reviews. *Information Systems Research*, Vol. 19, No. 4, December 2008, pp. 456–474.
- Li Y. M., Wu C. T., Lai C. Y. (2013). A social recommender mechanism for e-commerce: Combining similarity, trust, and relationship. *Decision Support Systems*, 55(3), pp. 740-752.
- Liang T. P., Lai H. J., Ku Y. C. (2007) Personalized content recommendation and user satisfaction: Theoretical synthesis and empirical findings. *Journal of Management Information Systems*, 23(3), pp. 45-70.
- Lim E.-P., Nguyen V.-A., Jindal N., Liu B., Lauw H.W. Detecting Product Review Spammers using Rating Behaviors. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2010)*.
- Liu B. (2010) Sentiment analysis and subjectivity. In N. Indurkha, F.J. Damerau (Eds.), *Handbook of Natural Language Processing* 2nd edition, 2010, pp. 1–38.

- Liu B. (2012) Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.
- Liu B., Cao Y., Lin C.-Y., Huang Y., Zhou (2007) Low-quality product review detection in opinion summarization. in Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), 2007, pp. 334-342.
- Liu F., Yu C., Meng W. (2004) Personalized Web search for improving retrieval effectiveness. IEEE Transaction of Knowledge and Data Engineering, 16, 2004, pp. 28–40.
- Liu X., Yang P., Fang H. (2014) EntEXPO: An Interactive Search System for Entity-Bearing Queries. Computer Science, Volume 8416, 2014, pp. 784-788.
- Liu Y., Jin J., Ji P., Harding J.A , Fung R.Y.K. (2013) Identifying helpful online reviews: A product designer’s perspective. Computer-Aided Design, 45 (2013), pp.180–194.
- Liu X., Chen F., Fang H., Wang M. (2014) Exploiting entity relationship for query expansion in enterprise search. Information Retrieval, 17(3), 265-294.
- Lloyd S. P. (1957) Least square quantization in PCM. Bell Telephone Laboratories Paper.
- Lu Y., Tsaparas P., Ntoulas A., Polanyi L. (2010) Exploiting social context for review quality prediction. In Proceedings of International World Wide Web Confernece (WWW-2010).
- Lynch JJG, Ariely D (2000) Wine online: Search costs affect competition on price, quality, and distribution. Marketing Science, 19(1), pp. 83–103.
- MacQueen J. B. (1967) Some Methods for classification and Analysis of Multivariate Observations. In Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 1. University of California Press, pp. 281–297.
- Manning C.D., Raghavan P., Schutze H. (2008) Introduction to Information Retrieval. Cambridge University Press, 32 Avenue of the Americas, New York, NY 10013-2473, USA.

- March S.T., Storey V.C. (2008) Design Science in the IS Discipline—Introduction to the Special Issue on Design Science Research. *MIS Quarterly* Vol. 32 No. 4, December 2008, pp. 725-730.
- March S.T., Smith G.F. (1995) Design and natural science research on information technology. *Decision Support Systems*, 15 (1995), pp. 251-266.
- Markus M. L., Majchrzak A., Gasser, L. (2002) A Design Theory for Systems that Support Emergent Knowledge Processes. *MIS Quarterly*, (26:3), September, 2002, pp. 179-212.
- Marshall B., McDonald Chen H., Chung W. (2004) EBizPort: Collecting and Analyzing Business Intelligence Information. *Journal of American Society Information Science and Technology*, 55, 10, pp. 873–891.
- Marron J. S. (2015) Distance-weighted discrimination. *Wiley Interdisciplinary Reviews: Computational Statistics*.
- McKinney EH, Yoos CJ (2010) Information about information: A taxonomy of views, *MIS Quarterly*. 34(2), pp. 329–344.
- Mecca G., Raunich S., Pappalardo A. (2007) A new algorithm for clustering search results. *Data & Knowledge Engineering*, Volume 62, Issue 3, September 2007, pp. 504-522.
- Melucci M., A (2008) Basis for Information Retrieval in Context. *ACM Transaction of Information Systems*, 26, 3, 2008, Article No 14.
- Micarelli A., Sciarrone F. (2004) Anatomy and empirical evaluation of an adaptive Web-based information filtering system. *User Modeling and User-Adaption Interface*, 14, 2004, pp. 159-200.
- Micarelli A., Gasparetti F., Sciarrone F., Gauch S. (2007) Personalized search on theWorldWideWeb. In: Brusilovsky, P., Kobsa, A., Nejdl,W. (eds.) *The AdaptiveWeb*, 1 edition, pp. 195–230, Springer, Berlin.

- Moe W. W. (2003) Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream. *Journal of Consumer Psychology* (13:1/2), 2003, pp. 29-39.
- Moghaddam S., Jamali M., Ester M. (2012) ETF: Extended Tensor Factorization Model for Personalizing Prediction of Review Helpfulness. *WSDM'12*, February 8–12, 2012, Seattle, Washington, USA.
- Montoyo A. Martínez-Barco P., Balahur A. (2012) Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decision Support Systems* 53 (2012), pp. 675–679.
- Mudambi S.M., Schuff D. (2010) What makes a helpful online review? A study of customer reviews On Amazon.com. In: C. Saunders (Ed.), *MIS Quarterly*, 34:1, 2010, pp. 185–200.
- Narver J.C., Slater S.F. (1990) The effect of a market orientation on business profitability, *The Journal of Marketing* 45 (4) (1990), pp. 20–36.
- Nassirtoussi A.K., Aghabozorg S., Wah R. Y., Ngo D.C.L. (2014) Text mining for market prediction: A systematic review. *Expert Systems with Applications*, Volume 41, Issue 16, 15 November 2014, pp. 7653–7670.
- Natsev A. P., Haubold, A., Tešić, J., Xie, L., & Yan, R. (2007, September). Semantic concept-based query expansion and re-ranking for multimedia retrieval. In *Proceedings of the 15th international conference on Multimedia* (pp. 991-1000).
- Ngo-Ye T.L., Sinha A.P. (2012) Analyzing Online Review Helpfulness Using a Regression Relief-Enhanced Text Mining Method. *ACM Transaction Management Information Systems*, Article 10 (July 2012), 20 pages.
- Ngo-Ye T.L., Sinha A.P. (2014) The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, 17 January 2014, available online.

- Nguyen D., Overwijk A., Hauff C., Trieschnigg D., Hiemstra D., De Jong F. (2008) WikiTranslate: query translation for cross-lingual information retrieval using only Wikipedia. In Lecture Notes in Computer Science. Cross-Language Evaluation Forum (CLEF 2008), pp. 58–65.
- Norman D. (1998) *The Design of Everyday Things*. Currency Doubleday, New York, 1988
- Ogilvie P., Voorhees E., Callan J. (2009) On the number of terms used in automatic query expansion. *Information Retrieval*, 12, 666–679 (2009).
- Oliver R.L. (1995) Attribute need satisfaction in product usage satisfaction. *Psychology & Marketing*, 12(1), pp. 1–17.
- O'Mahony M.P., Smyth B. (2009) Learning to recommend helpful hotel reviews. In *Proceedings of the third ACM conference on Recommender systems*, 2009.
- O'Mahony M.P., Smyth B. (2010) A classification-based review recommender, in: M. Bramer (Ed.), *Research and Development in Intelligent Systems XXVI*, Springer-Verlag London Limited, London, 2010, pp. 49–62.
- Orlikowski W. J., Iacono C. S. (2001) Research Commentary: Desperately Seeking the 'IT' in IT Research A Call to Theorizing the IT Artifact. *Information Systems Research*, (12:2), June 2001, pp. 121-134.
- Osiński S., Stefanowski J., Weiss D. (2004) Lingo: Search Results Clustering Algorithm Based on Singular Value Decomposition. *Intelligent Information Processing and Web Mining Advances in Soft Computing Volume 25*, 2004, pp 359-368.
- Osiński S., Weiss D. (2005) A concept-driven algorithm for clustering search results. *Intelligent System*, 20 (2005), pp. 48-54.
- Ott M., Choi Y., Cardie C. (2011) Hancock J.T. Finding deceptive opinion spam by any stretch of the imagination. in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011)*.

- Otterbacher J. (2009) Helpfulness in online communities: a measure of message quality. Proceedings of the 27th international conference on human factors in computing systems, Boston (MA, USA), ACM, 2009, pp. 955–64.
- Page L., Brin S., Motwani R., T Winograd T. (1999) The PageRank Citation Ranking: Bringing Order to the Web.
- Palmer J.W., Web Site Usability, Design, and Performance Metrics, Information Systems Research, Volume 13, Issue 2, June 1, 2002, 151-167.
- Pana Z., Xinge Youa, Hong Chenb, Dacheng Taoc, Baochuan Panga (2013) Generalization performance of magnitude-preserving semi-supervised ranking with graph-based regularization. Information Sciences, Volume 221, 1 February 2013, pp. 284–296.
- Pang B., Lee L. (2008) Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, Vol. 2, Nos. 1-2 (2008), pp. 1-35.
- Pang B., Lee L., Vaithyanathan S. (2002) Thumbs up? Sentiment classification using machine learning techniques, Proceedings of EMNLP 2002, 2002, pp. 79–86.
- Parboteeah, D. V., Valacich J. S., and Wells J. D. (2009) The Influence of Website Characteristics on a Consumer's Urge to Buy Impulsively. Information Systems Research, (20:1), pp. 60-78.
- Park D.H., Kim S. (2008) The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. Electronic Commerce Research and Applications 7 (2008), pp. 399–410.
- Park L. A., Ramamohanarao K. (2007). Query expansion using a collection dependent probabilistic latent semantic thesaurus. In Advances in Knowledge Discovery and Data Mining (pp. 224-235). Springer Berlin Heidelberg.
- Pavlou P. A., Liang H., Xue Y. (2007) Understanding and mitigating uncertainty in online exchange relationships: a principal-agent perspective. MIS Quarterly, 31, 1, 2007, pp. 105–136.

- Peffer K., Tuunanen T., Rothenberger M.A., Chatterjee S. (2007) A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, Volume 24 Issue 3, Winter 2007-8, pp. 45-78.
- Piramuthu S., Kapoor G., Zhou W., Mauw S. (2012) Input online review data and related bias in recommender systems. *Decision Support Systems*, Volume 53, Issue 3, June 2012, pp. 418-424.
- Pitkow J., Schutze H., Cass T., Cooley R., Turnbull D., Edmonds A., Adar E., Breuel T. (2002) Personalized search. *Communication of ACM*, 45, 2002, pp. 50–55.
- Porcel C., Tejada-Lorente A., Martínez M.A., E. Herrera-Viedma E. (2012) A hybrid recommender system for the selective dissemination of research resources in a Technology Transfer Office. *Information Science*, 184 (2012), pp. 1-19.
- Quinlan J. R. (1979) *Discovering rules by induction from large collections of examples. Expert Systems in the micro electronic age*, D Michie (Ed.), Edinburgh University Press, pp. 168-201.
- Quinlan J. R. (1986) *Simplifying Decision Trees*, *International Journal of Man-Machine Studies*, 27, pp. 221-234.
- Quinlan J. R. (1993) *C4.5: Programs for Machine Learning*. Morgan Kaufman, San Mateo, CA.
- Ramage D., Heymann P., Manning C.D., Garcia-Molina H. (2009) *Clustering the Tagged Web. Proceeding of WSDM '09 Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pp. 54-63.
- Rattenbury T., Naaman M. (2009) *Methods for extracting place semantics from Flickr tags. ACM Transactions on the Web* 3, 1 (January 2009), Article 1.
- Rendle S. (Steffen) (2012) *Factorization machines with libFM. ACM Transactions on Intelligent Systems and Technology* 3, 3 (May 2012), Article 57.

- Ricci F., Rokach L., Shapira B. (2011) Introduction to recommender systems handbook, Springer US, pp. 1-35.
- Riezler S., Vasserman A., Tsochantaridis I., Mittal V., Liu Y. (2007, June) Statistical machine translation for query expansion in answer retrieval. In ANNUAL MEETING-ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, Vol. 45, No. 1, p. 464.
- Riloff E., Patwardhan S., and J. Wiebe J. (2006) Feature subsumption for opinion analysis. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2006.
- Rubinstein R. Y., Kroese D. P. (2004) The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning. Information Science and Statistics, Springer, 2004.
- Sagarna R., Mendiburu A., Inza I., Lozano J.A. (2014) Assisting in search heuristics selection through multidimensional supervised classification: A case study on software testing. Information Sciences, 258 (2014), pp. 122–139.
- Sahami Mehran, Heilman T.D. (2006) A web-based kernel function for measuring the similarity of short text snippets. In Proceedings of the 15th international conference on World Wide Web, pp. 377-386.
- Sakai T., M. Manabe, M. Koyama (2005) Flexible pseudo-relevance feedback via selective sampling, ACM Transaction of Information Systems, 4, 2, pp. 111–35.
- Salton G., Fox E.A. and Wu H. (1983) Extended Boolean information retrieval. Communications of the ACM, Vol. 36(11), Nov. 1983, pp. 1022-1036.
- Salton G., McGill M. (1983), Introduction to Modern Information Retrieval, McGraw Hill, 1983, New York, NY.
- Sánchez, J., Perronnin, F., Mensink, T., & Verbeek, J. (2013). Image classification with the Fisher vector: Theory and practice. International journal of computer vision, 105(3), 222-245.

- Sanjay K. Arora, Alan L. Porter, Jan Youtie, Philip Shapira, Capturing new developments in an emerging technology: an updated search strategy for identifying nanotechnology research outputs, *Scientometrics*, April 2013, Volume 95, Issue 1, pp 351-370.
- Santos R.L.T., Macdonald C., Ounis I. (2010) Exploiting query reformulations for Web search result diversification. *Proceeding of 19th International Conference on World Wide Web (WWW 2010)*, pp. 881–890. ACM, Raleigh (2010).
- Savoy, J. (2003). Cross-language information retrieval: experiments based on CLEF 2000 corpora. *Information processing & management*, 39 (1), 75-115.
- Sebastiani F. (2002) Machine learning in automated text categorization. *ACM Computing Surveys*, vol. 34, pp. 1-47.
- Segura A., Christian Vidal-Castro, Mateus Ferreira-Satler, Salvador-Sánchez (2014) Domain Ontology-Based Query Expansion: Relationships Types-Centered Analysis Using Gene Ontology. *Advances in Intelligent Systems and Computing Volume 232*, 2014, pp. 183-188.
- Senecal S., J. Nantel (2004) The influence of online product recommendations on consumers' online choices, *Journal of Retailing* 80 (2) (2004) 159–169.
- Shani G., A. Gunawardana, Evaluating recommendation systems, in: F. Ricci, L. Rokach, B. Shapira, P.B. Kantor (Eds.), *Recommendation System Handbook*, Springer US, 2011, 257–297.
- Shi Y., Martha Larson M., and Alan Hanjalic A. (2010) Towards understanding the challenges facing effective trust-aware recommendation. In *Proceedings of Workshop on Recommender Systems and the Social Web*, ACM, New York, NY, 2010.
- Shiri A. A., Revie C. (2003). The effects of topic complexity and familiarity on cognitive and physical moves in a thesaurus-enhanced search environment. *Journal of Information Science*, 29(6), 517-526.

- Shiv B., Fedorikhin A. (1999) Heart and Mind in Conflict: the Interplay of Affect and Cognition in Consumer Decision Making. *Journal of Consumer Research*, Vol. 26, No. 3 (December 1999), pp. 278-292.
- Simon H. A. (1996) *The Sciences of the Artificial*. The MIT Press, Cambridge, Massachusetts, USA.
- Singhal A. (2001). Modern information retrieval: A brief overview. *IEEE Data Engineering Bulletin*, 24(4), pp. 35-43.
- Smith S.P., Johnston R.B., Howard S. (2011) Putting Yourself in the Picture: An Evaluation of Virtual Model Technology as an Online Shopping Tool. *Information Systems Research*, 22(3), pp. 640-659.
- Smyth, B., Balfe, E. (2006) Anonymous personalization in collaborative Web search. *Information Retrieval*, 9, 2006, pp.165-190.
- Song D., R.Y.K. Lau, P.D. Bruza, K.F. Wong, D.Y. Chen, An adaptive information agent for document title classification and filtering in document-intensive domains, *Decision Support Systems*, 44 (1) (2008) 251–265.
- Song J.H., Zinkhan G. M. (2008) Determinants of Perceived Web Site Interactivity. *Journal of Marketing* Vol. 72 (March 2008), pp. 99–113.
- Song M., Song I. Y., Allen R. B., Obradovic Z. (2006) Keyphrase extraction-based query expansion in digital libraries. In *Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries*, pp. 202-209.
- Spink A., Wolfram D., Jansen M. B., Saracevic T. (2001). Searching the web: The public and their queries. *Journal of the American society for information science and technology*, 52(3), 226-234.
- Steichen B., Lawless S., O’connor A., Wade V. (2009) Dynamic hypertext generation for reusing open corpus content. In: *20th ACM Conference on Hypertext and Hypermedia (Hypertext 2009)*, pp. 119–128.

- Steinhaus H. (1957) Sur la division des corps matériels en parties. Bull. Academic Polonium Science. (in French) 4 (12), pp. 801–804.
- Sugiyama K., Hatano K., Yoshikawa M. (2004) Adaptive Web search based on user profile constructed without any effort from users. In Proceeding of 13th International Conference on World Wide Web (WWW 2004), ACM, New York (2004), pp. 675–684.
- Sun H., Fang Y., Hsieh J. P. A. (2014) Consuming information systems: An economic model of user satisfaction. Decision support systems, 57, pp. 188-199.
- Sun R., Ong C. H., Chua T. S. (2006). Mining dependency relations for query expansion in passage retrieval. In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 382-389).
- Swaab R.I., Postmes T., Neijens P., Kiers M.H., Dumay A.C.M. (2002) Multi-party negotiation support: Visualization and the development of shared mental models. Journal of Management Information Systems, 19, 1 (2002), pp. 129-150.
- Tan B., Peng F. (2008). Unsupervised query segmentation using generative language models and wikipedia. In Proceedings of the 17th international conference on World Wide Web (pp. 347-356).
- Tanweer M. R., Suresh S., Sundararajan N. (2015). Self-regulating particle swarm optimization algorithm. Information Sciences, 294, pp. 182-202.
- Taylor H., Dillon S., Van Wingen M. (2010) Focus and Diversity in Information Systems Research: Meeting the Dual Demands of A Healthy Applied Discipline. MIS Quarterly, Vol. 34 No. 4, December 2010, pp. 647-667.
- Teevan J., Collins-Thompson K., White R. W., Dumais S. T., Kim Y. (2013) Slow Search: Information Retrieval without Time Constraints. In Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval, p. 1.
- Tegegne T., Weide Th.P. (2014) Enriching queries with user preferences in healthcare. Information Processing & Management, Volume 50, Issue 4, July 2014, pp. 599-620.

- Terveen L., Hill W., Amento B., McDonald D., Creter J. (1997) PHOAKS: a system for sharing recommendations. *Communications of the Association for Computing Machinery (CACM)* 40 (3) (1997), pp. 59–62.
- Theobald M., Schenkel R., Weikum G. (2005) Efficient and self-tuning incremental query expansion for top-k query processing. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 242-249). ACM.
- Torkestani J.A., An adaptive learning to rank algorithm: Learning automata approach, *Decision Support Systems* 54 (2012) 574–583.
- Tsaparas P., Ntoulas A., Terzi E. (2011) Selecting a Comprehensive Set of Reviews. in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD-2011)*.
- Tsur O., Rappoport A. (2009) Revrnk: A fully unsupervised algorithm for
- Turney Y. P. 2001. Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In *Proceedings of the 12th European Conference on Machine Learning*.
- Turney P.D. (1995) Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm, *Journal of Artificial Intelligence Research*, 2, pp. 369-409.
- Turney P., Littman M. L., Bigham J., Shnayder V. (2003). Combining independent modules to solve multiple-choice synonym and analogy problems.
- Vaishnavi V., Kuechler B. (2011) Design Science Research in Information Systems. <http://desrist.org/design-research-in-information-systems/> (September 30, 2011).
- Van der Heijden H., Verhagen T. (2004) Online store image: conceptual foundations and empirical measurement. *Information & Management*, 41, 5, 2004, pp. 609–617.

- van Dijck J. (2009) Users like you? Theorizing agency in user-generated content. *Media, Culture & Society*, 31 (2009), pp. 41-58.
- Vapnik V. (1995) *The Nature of Statistical Learning Theory*. Springer, N.Y., 1995. ISBN 0-387-94559-8.
- Vapnik V. N., Vapnik V. (1998) *Statistical learning theory (Vol. 1)* New York: Wiley.
- Walls J. G., Widemeyer G. R., El Sawy O. A. (1992) Building an Information System Design theory for Vigilant EIS. *Information Systems Research*, (3)1, pp. 36-59.
- Walls J., Widmeyer G., El Sawy, O. (2004) Assessing Information System Design Theory in Perspective: How Useful was our 1992 Initial Rendition? *Journal of Information Technology Theory & Application (JITTA)*, 6, 2 (2004), pp. 43-58.
- Wang G., Xie S., Liu B., Yu P.S. Identify Online Store Review Spammers via Social Review Graph. *ACM Transactions on Intelligent Systems and Technology*, Accepted for publication, 2011.
- Wang G.A., Jiao J., Abrahams A.S., Fan W., Zhang Z. (2013) ExpertRank: a topic-aware expert finding algorithm for online knowledge communities. *Decision Support Systems* (54:3) (2013), pp. 1442–1451.
- Wang R.Y., Strong D.M. (1996) Beyond accuracy: what data quality means to data consumers, *Journal of Management Information Systems* 12 (4) (1996) 5–33.
- Wang W, Benbasat I. (2009) Interactive decision aids for consumer decision making in e-commerce: The influence of perceived strategy restrictiveness. *MIS Quarterly*, 33(2):293–320.
- Wang W., Benbasat I. (2014) Research Note—A Contingency Approach to Investigating the Effects of User-System Interaction Modes of Online Decision Aids. *Information Systems Research* 24(3), pp. 861-876.

- Witten I. H., W. P. Gordon, E. Frank, C. Gutwin, C. G. Nevill-Manning, KEA: Practical Automatic Keyphrase Extraction, DL '99 Proceedings of the fourth ACM conference on Digital libraries, 254-255, ACM New York, NY, USA, 1999.
- Wolfenbarger, M., Gilly, M. C. (2001) Shopping online for freedom, control, and fun. *California Management Review*, 43, 2, 2001.
- Wu I.-C., Lin Y.S. (2012) WNavis: Navigating Wikipedia semantically with an SNA-based summarization technique. *Decision Support Systems*, Volume 54, Issue 1, December 2012, pp. 46–62.
- Xiong L., Xi Chen, Tzu-Kuo Huang, Jeff Schneider, and Jaime G. Carbonell (2010) Temporal collaborative filtering with Bayesian probabilistic tensor factorization. *Proceedings of SIAM Data Mining 2010 (SDM'10)*, pp. 211–222.
- Xu J., Benbasat I., Cenfetelli R.T. (2014) The Nature and Consequences of Trade-off Transparency in the Context of Recommendation Agents. *MIS Quarterly* Vol. 38 No. 2, pp. 379-406/June 2014.
- Xu J.D., Izak Benbasat I., Cenfetelli R. T. (2014) Research Note – The Influences of Online Service Technologies and Task Complexity on Efficiency and Personalization. *Information Systems Research* 25(2):420-436
- Xu Y., Jones G. J. F., Wang B. (2009) Query dependent pseudo-relevance feedback based on wikipedia. In *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM Press, 59–66.
- Yin Z., Shokouhi M., Craswell N. (2009) Query expansion using external evidence. *Lecture notes in computer science*. In: *31st European Conference on Information Retrieval (ECIR 2009)*, pp. 362-374. Springer, Toulouse (2009).
- Yin Z., Shokouhi M., Craswell N. (2009). Query expansion using external evidence. In *Advances in Information Retrieval*, pp. 362-374. Springer Berlin Heidelberg.
- Zamir O. (Oren), Etzioni O. (1999) Grouper: a dynamic clustering interface to Web search results. *Computer Networks*, Volume 31, Issues 11–16, 17 May 1999, pp. 1361-1374.

Zhang Z. Varadarajan B. (2006) Utility Scoring of Product Reviews. Proceedings of the 15th ACM international Conference on Information and Knowledge Management, 2006, pp. 51–57.

Zhang P. (2000) The Effects of Animation on Information Seeking Performance on the World Wide Web: Securing Attention or Interfering with Primary Tasks. Journal of the Association for Information Systems (1), 2000, Article 1.

Zhou D., Lawless S., Wade V.: Improving search via personalized query expansion using social media. Information Retrieval, 2012, pp. 1-25.

APPENDIX I

K-Means Clustering Overview

In chapter 5, section 5.1.2, we presented our proposed Directed Particle Swarm Optimization and K-Means (DPSO-KM) algorithm, which utilizes K-Means technique to produce the relevance set from a review pool. K-Means clustering is the most popular technique used in information retrieval and extraction as well as web personalization (Cagnina et al., 2014). In this appendix, we provide a brief overview of the classic K-Means technique.

The term K-Means is regarded to be first coined by MacQueen (1967). However, a preliminary K-Means algorithm was described in Steinhaus (1957). And, Lloyd (1957) is considered as the first contribution to K-Means method (Amorim & Mirkin, 2012). In machine learning and information retrieval field, the work of Lloyd (1957) is associated with the work of Forgy (1965) by the so-called Lloyd-Forgy algorithm. It reflects the fact that Forgy (1965) proposed a K-Means algorithm similar to the algorithm proposed by Lloyd (1957). Although other K-Means algorithms have been proposed, they are still in the same vein with MacQueen's algorithm or Lloyd-Forgy's algorithm (Dasgupta & Freund 2009). The difference between those two algorithms is mainly that former provides a cost function while the latter does not use it.

So far, the K-Means algorithm proposed by MacQueen (1967) has been the most applied technique in information retrieval and web personalized (Cagnina et al., 2014; Jain, 2010). Since

the core function of MacQueen’s algorithm is basically the same as that of Lloyd-Forgy’s algorithm. Hereinafter, we focus on MacQueen’s algorithm.

The procedure of MacQueen’s algorithm is straightforward, which first selects K webpages as the initial centroids. They will be used to construct the first set of clusters. Then, the algorithm moves on to calculate the distances between the initial centroids and each webpage in an iterative manner. A webpage is assigned to the nearest centroid (Figure 20 (a)).

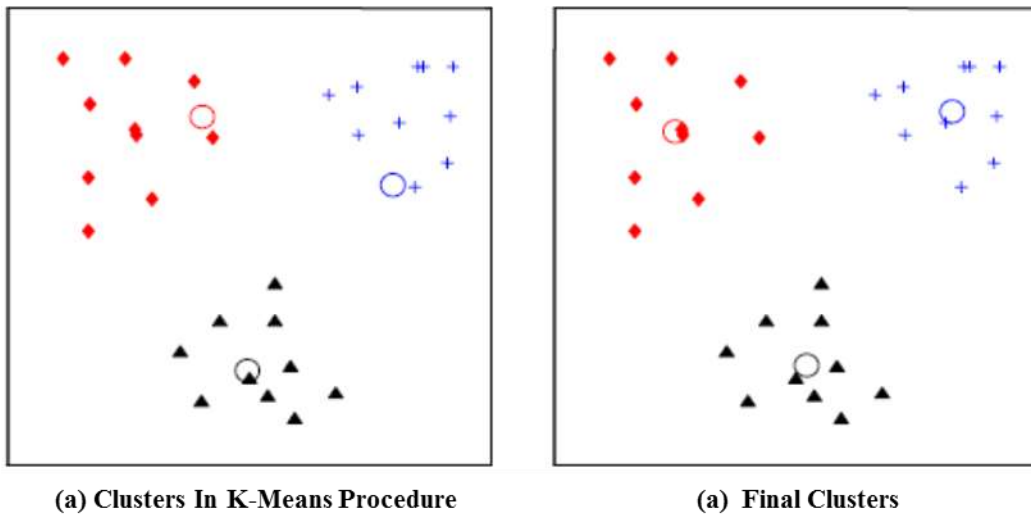


Figure 20. K-Means Results from Figure 4 of Jain (2010)

Then, the initial centroids are updated by using the cost function (CF):

$$\Delta c_{\min}(t) = \theta(t)[w(t) - c_{\min}(t-1)] \quad (\text{CF})$$

In CF, $c_{\min}(t)$ is the nearest centroid; $w(t)$ is the document/webpage; t is the time; $\theta(t)$ is the adaption rate. The initial centroids are updated by the adaption rate, the webpage $w(t)$, and

the nearest centroid $c_{\min}(t-1)$. The centroids are typically updated many times. After a number of updating, the final clustering result is produced (Figure 20 (b)), which meets the minimum cost.

MacQueen (1967) provided the function for adaption rate: $\theta(t) = 1/n_t$ where n_t is the size of the cluster that is the nearest one to the centroid. In the literature, there are various complex adaption rate functions proposed after MacQueen (1967), yet MacQueen's function has demonstrated high effectiveness (Jain, 2010).

APPENDIX II

Support Vector Machine Overview

In chapter 2, we discussed that web personalization and review helpfulness research most often uses Support Vector Machine (SVM) to identify useful webpages or online reviews for consumers. In this appendix, we discuss the common classification SVM algorithm applied in web personalization and review helpfulness research.

A large body of SVM algorithms has been proposed by the researchers in many research fields. The proposed algorithms have their core function in common, which was introduced in the seminal work Vladimir & Vapnik (1998). In their work, Vladimir and Vapnik proposed the SVM function for statistical learning (Vladimir & Vapnik 1998). Their proposal refined the early SVM works, e.g., Vapnik (1995), but remained the basic promise offered by the early works.

Since late 1990's, the SVM has been successfully applied in various classification problems such as document classification, pattern recognition, as well as information retrieval and extraction from the Web. The SVM has also become the most-applied technique in utility-sensitive review analysis (refer to section 2.3) since the SVM is capable to handle a huge dimensional vector spaces and achieve better accuracy comparing to other statistical learning methods (Marron, 2015).

In general, a SVM algorithm includes a generalized linear model or a kernel function (Marron, 2015) in a high dimensional vector space, which is trained by an optimization

algorithm. The optimization accounts the learning bias involved in the process. The algorithm is attempted to choose a hyper plane space that can maximum the margin. The functions of the SVM are as following.

$$\text{If } Y_j = +1, wx_j + b \geq 1 \quad [1]$$

$$\text{If } Y_j = -1, wx_j + b \leq -1 \quad [2]$$

$$\text{For all } j, y_j (wx_j + b) \geq 1 \quad [3]$$

In [1], [2], and [3], x_j is j^{th} vector, and w is the weight. The three equations are enforced automatically by a SVM software packages. The SVM algorithm searches the optimal hyper plane from the hyper plane space. The optimal hyper plane should maximize the margin.

The SVM procedure results in a classification with the maximum space between the boundaries of the hyper plan. The resultant classification can achieve very high accuracy when the size of data is large. A desirable classification is identified when the resultant hyper plane is at the farthest distance from the data regardless their individual positions. In such case, the hyper plane bisects the lines between the closest data points on the convex boundaries (Figure 21).

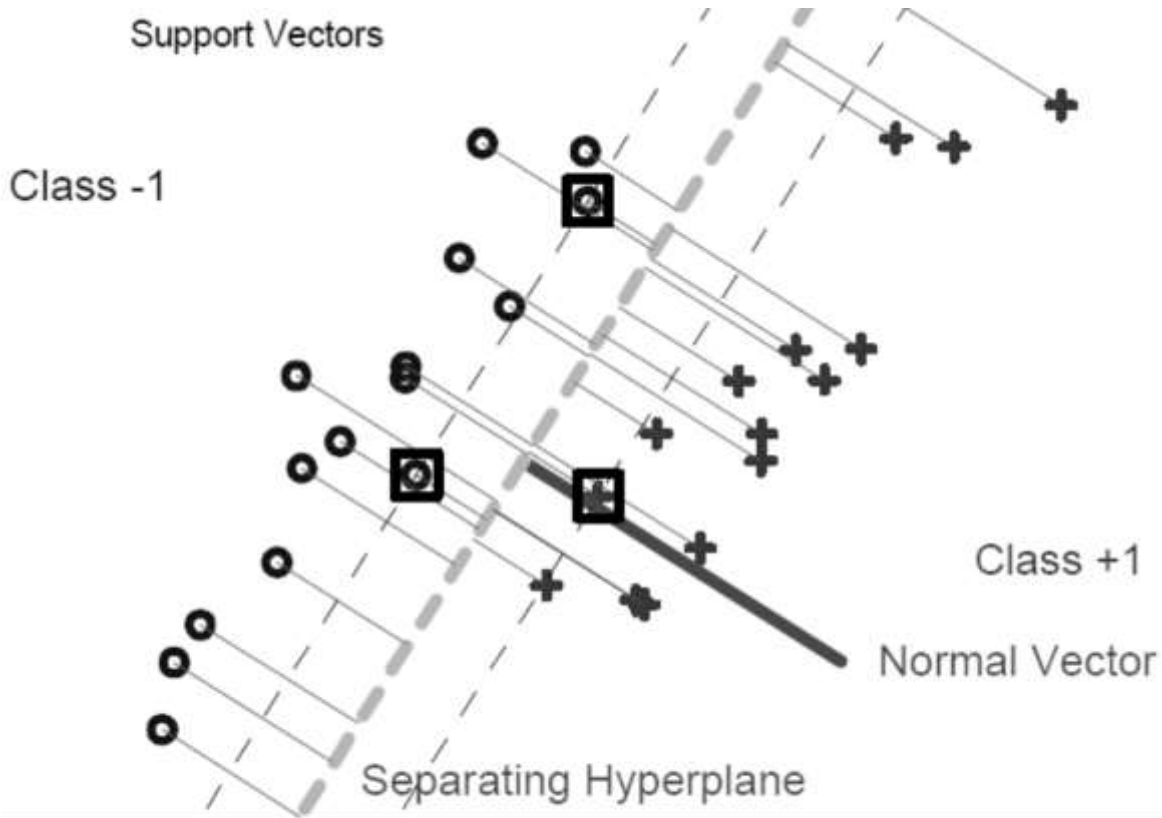


Figure 21. Support Vector Machine from Figure 1 of Marron (2015)

Moreover, the distance of the closest data point on the hyper plane to origin is calculated by maximizing x . In fact, the SVM calculates the distances of the closest data points on both sides to origin. Then, the algorithm calculates the aggregated distance from the hyper plane to the nearest points by solving the distances. The maximum margin is calculated by $2 / \|w\|$.

In addition, the SVM algorithm calculates w and b by using the Lagrange's multiplier α_j . The function is as following:

$$w = \sum \alpha_j x_j$$

$\mathbf{b} = y_n - \mathbf{w}x_n$ for any x_n when $\alpha_n \neq 0$

The SVM classifies the webpages with $f(x) = \sum \alpha_j y_j x_j * x + \mathbf{b}$, which can produce optimal classification result.

APPENDIX III

Decision Tree Overview

In chapter 2, we mentioned that some web personalization and review helpfulness research utilizes Decision Tree (DT) model to predict the utility of a webpage or an online review. Prior research has proposed a large number of DT algorithms using different partition techniques. This appendix provides an overview of the major DT techniques.

The early most influential works in DT techniques included Quinlan (1979 & 1986). These works proposed the ID3 approach, which produces a decision tree with nodes and branches. A node without any subsequent branch is called a leaf node or leaf. The ID3 utilized a greedy top-down technique where the feature A was selected as the root node. Then, the training data were separated into different subsets by the feature A. For each subset, the same process was applied to further split the subset into smaller sets. The number of the leaf nodes was determined. The ID3 used QF1 to calculate the expected entropy E for A.

$$E(A) = \sum_{a \in A} P(a) \sum_{c \in C} -P(a|c) \log_2(P(a|c)) \quad \text{QF1}$$

In QF1, A is the feature A; C is the class. Furthermore, the ID3 calculated the $I_A = E(D) - E(A)$ to evaluate information gain. The E(D) was the entropy before splitting. Also, the ID3 approach used predictive accuracy to evaluate the quality of the produced decision tree. The algorithm attempted to maximize the overall accuracy.

To correct the bias introduced by the ID3 procedure, Quinlan (1993) proposed C4.5 machine learning algorithm, the GINI, which could calculate the gain ratio as following.

$$GainRatio_A = \frac{I_A}{Info_A} \text{ where } Info_A = \sum_{a \in A} -\frac{N_a}{N} \log_2 \frac{N_a}{N}$$

Quinlan (1986) introduced two post-pruning methods to reduce misclassification. The first pruning used a test set and classified the original tree T. Suppose S is a sub-tree of T. The pruning algorithm replaced S with the best possible leaf. If the new misclassification was equal to or less than that produced by T and S, then S was replaced by the leaf. The second pruning used pessimistic technique. Suppose the sub-tree S has L(S) leaves, K cases in a leaf, and J misclassified cases. If we replace S by the best leaf, E cases are misclassified. The pessimistic pruning replaced S by the best leaf when $E + \frac{1}{2}$ within one standard error of $\sqrt{J + L(S)/2}$.

A popular cost-base DT algorithm was proposed by Turney (1995), which was called ICET (Inexpensive Classification with Expensive Tests). It used genetic methods for cost. ICET system first produced initial decision trees. Then, the algorithm evaluated the trees by using the Fitness Function that combined initial tree to produce a new set of trees repeatedly until the threshold was met. The system utilized the following cost function:

$$ICF_A = 2^{InfoGain_A} - 1 / (C_A + 1)^\omega$$

Each example had the parameters of C_A , ω , and CF (C_A and ω are bias). The CF was the degree of pruning. ICET first divided the training data into two equal groups: 1) a training set and a testing set. An initial tree was derived from the sub-training set where the examples. Then, the cost function was used to calculate the average cost of the classifications. Next, ICET

generated the next tree by using the roulette wheel selection scheme. It selected trees using the probability corresponding to their fitness. A threshold number was used to determine when to stop the evaluation. Then, the best fitted tree was selected by using the fitness function. Turney (1995) used a set of non-greedy algorithms to demonstrate the benefits of the ICET.

Another seminal work was Freund & Schapire (1997). They proposed the well-accepted boosting algorithm called AdaBoost. In Figure 22, we use the Figure 1 in Freund & Schapire (1997) to present the AdaBoost algorithm where $(x_1, y_1), \dots, (x_m, y_m)$ were input; x_i was an item of domain X ; y_i was an item of domain $Y = \{-1, +1\}$. AdaBoost utilized a learner in t runs ($t = 1, \dots, T$). The weights were used in the training to derive the classifier $h_t: X \rightarrow \mathbb{R}$, which was evaluated by the error. $H(x)$ in Figure 22 is the final tree presented by the algorithm.

Given: $(x_1, y_1), \dots, (x_m, y_m)$ where $x_i \in X, y_i \in Y = \{-1, +1\}$
 Initialize $D_1(i) = 1/m$.
 For $t = 1, \dots, T$:

- Train base learner using distribution D_t .
- Get base classifier $h_t: X \rightarrow \mathbb{R}$.
- Choose $\alpha_t \in \mathbb{R}$.
- Update:

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution).

Output the final classifier:

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right).$$

Figure 22. AdaBoost

Recent improvement of DT algorithms includes the popular work, Barros et al. (2011). They proposed the E-Motion decision tree algorithm that derives multiple initial trees on the basis of the shapes and sizes of the nodes. Then, the algorithm pruned each initial tree by combining single nodes. The combination used the expected standard deviation reductions (SDR) calculated as following:

$$SDR = sd(D) - \sum_{i=1}^k \frac{|D_i|}{|D|} \times sd(D_i)$$

In the function, $sd(D)$ was the standard deviation, D was the set of the examples, $|D|$ was the size of the node, D_i was the set of the examples, and $|D_i|$ was the size of a partition.

The initial trees were optimized at the later stages. The E-Motion optimized the trees using weighted function of root mean squared error, mean absolute error, and tree size. Alternatively, a lexicographic analysis was used for the optimization. Moreover, the E-Motion used two different strategies to optimize the initial tree. The first strategy was the shrinking optimization where a subtree was replaced by a leaf node. The second strategy was the expanding optimization where a leaf node was replaced by a two-level subtree. The algorithm used a set of thresholds to determine which strategy was used at a leaf node. Finally, a filter was applied to guarantee consistency of the models at each leaf node.

More recently, Bina et al. (2013) proposed a Decision Tree Forest (DTF) method, which used DT algorithm to drive a decision forest. Then, each branch of a decision tree was

transformed into a Markov Logic Networks (MLNs). Logistic regression was used to produce the weight for a MLN. Then, the data independence was calculated by the following function:

$$P(T, J_1, \dots, J_m | c) = P(\mathbf{a} | c) \prod_{i=1}^m \prod_{r=1}^{rows_i} P(J_{i,r} | c)$$

In the function above, c was the MLN, \mathbf{a} was an input, and J_i was a join. A multi-relational classification model as followed was proposed to classify the data.

$$\frac{P(c(t) = 0 | T, J_1, \dots, J_m)}{P(c(t) = 1 | T, J_1, \dots, J_m)} = \frac{P(T, J_1, \dots, J_m | c(t) = 0) P(c(t) = 0)}{P(T, J_1, \dots, J_m | c(t) = 1) P(c(t) = 1)}$$

Furthermore, the researchers proposed an algorithm to iteratively use the multi-relational classification model. The empirical evaluation supported that the proposed method could provide more accurate prediction than the previous method.