APPROVAL SHEET

Title of Dissertation: Product Feature Driven Personalization of Online Consumer Reviews

Name of Candidate: Anupama Dash
Doctor of Philosophy, 2018

Dissertation and Abstract Approved: (*Signature of Supervising Professor)
(DONGSONG ZHANG)
(Professor)
(Information Systems)

Date Approved: _____

NOTE: *The Approval Sheet with the original signature must accompany the thesis or dissertation. No terminal punctuation is to be used.

ABSTRACT

Title of Document:	PRODUCT FEATURE DRIVEN PERSONALIZTION OF ONLINE CONSUMER REVIEWS
	Anupama Dash, Ph.D, May 2018
Directed By:	Professor Dongsong Zhang, Ph.D. Department of Information Systems University of Maryland Baltimore County

Many consumers rely on online product reviews for learning other consumers' opinions in favor of or against purchasing a product. With many people writing product reviews, there is an overwhelming number of reviews available online for a consumer to go through, causing navigation of those reviews tedious and ineffective. More importantly, despite that consumers have different information preferences for certain products of their interest, the existing online review platforms do not provide a personalized presentation of product reviews based on consumers' product feature preferences. There is a lack of theoretical frameworks for building personalized review rankings that can help consumers locate relevant and helpful reviews efficiently. Also, there is little empirical evidence available to demonstrate the effectiveness of personalizing review ranking systems.

To address these issues, this research proposes, implements, and evaluates a product-feature driven framework to personalize the presentation of online product reviews. The proposed product feature driven personalization of online product reviews (FDPPR) framework presents product reviews based on a consumer's product feature preferences. The research involves characterizing a large number of product reviews using natural language processing techniques. A latent class regression (LCR) model is then developed to present a unique personalized order of reviews to a prospective consumer based on his/her product feature preferences. In addition, an online user study is conducted to evaluate the performance of the proposed framework. The results show that the participants find the reviews clustered and presented by FDPPR to be relevant and satisfactory and provide better knowledge about a product than the baseline online review platform.

The major contributions of this research are (1) predicting helpfulness of product reviews based on individual product features, (2) reducing information overload for prospective consumers by providing personalized review ranking, and (3) developing a user evaluation methodology to measure the effectiveness of the personalization framework.

The practical implications of the research include (1) helping retailers present the most relevant reviews to the consumers first, (2) assisting consumers to quickly get the essence of a lot of reviews based on their product feature preferences, and (3) helping manufacturers of the products understand consumers' expressed needs.

PRODUCT FEATURE DRIVEN PERSONALIZATION OF ONLINE CONSUMER REVIEWS

By

Anupama Dash

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore County, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2018

Acknowledgements

I am deeply indebted to Professor Dr. Dongsong Zhang for his encouragement, advice, mentoring, and research support throughout my master's and doctoral studies. I am especially grateful to him for giving me the freedom to explore different areas of research and his guidance in helping me understand the different problems in product reviews recommendation. Also, I am very grateful to Dr. Lina Zhou for her generous guidance and inspiration throughout my research.

I would also like to thank my other dissertation committee members who distinguish themselves through both brilliant scholarship and exceptional teaching: Dr. Gunes Koru, Dr. Zhiyuan Chen, and Dr. Siddharth Kaza (Towson University), for their help, encouragement, and valuable feedback in this work; without their contributions it would not have been possible.

I am indebted to my good friends whose understanding, example, and encouragement inspire me to achieve higher goals. I truly hold dear their friendships. I would also like to gratefully acknowledge my elementary and secondary school teachers for providing me the sound foundation upon which I could realize my dream.

A very special thank go to my relatives for their comfort and encouragement. I owe special thanks to my parents, parents-in-law, my husband Jyotirmaya and my little and lovely daughter Aarima. They had made me believe that I can do anything I want and never stop having faith in me. Their unconditional love and support throughout my life is immeasurable, and I am forever indebted to that.

Last, but not the least, thank you GOD for everything.

TABLE OF CONTENTS

List of Tables

List of Figures

Chapter 1. Introduction

As consumers are increasingly using the internet to communicate with each other, consumer opinion, a common form of electronic word-of-mouth (eWOM), has gained a lot of importance [1]. One of the important channels to spread consumer opinions, is online review websites such as Amazon.com and Yelp.com that provide an online platform for consumers to write product reviews [2]. Online review platforms (ORPs) enable consumers to express their feelings, experiences, and opinions about a product and/or a service in a textual format that are accessible to others, including prospective consumers. For a given product, typical ORPs contain detailed product specifications and reviews from reviewers. After a consumer purchases and uses a product, he/she may write a review describing his/her experience with the product features. The reviewer also provides an overall star rating for the product. A consumer can read other people's reviews and provide a helpfulness rating for one or more reviews. A prospective consumer reads the reviews of a product/service of his/her interest to

Table 1. Terminology

Reviewer	A consumer who writes product/service reviews
Prospective Consumer	A consumer who is planning to buy a product/service
Online Review Platform (ORP)	A platform that enables reviewers to write reviews and facilitates prospective consumers to read the reviews
Specified Preference	The product features selected by prospective consumers when searching for a product
Product Features	Technical features of a product (e.g. storage space in a digital camera, remote of a TV or keyboard in a laptop etc.)
Review Stylistic Characteristics	Textual, meta-characteristics of the review (e.g. length of review, star rating)
Review Sentiment	The overall positive or negative polarity of a review (e.g. positive/negative score of a review)

form a judgment regarding the product/service. An overview of ORPs and the typical workflow is shown in Figure 1.

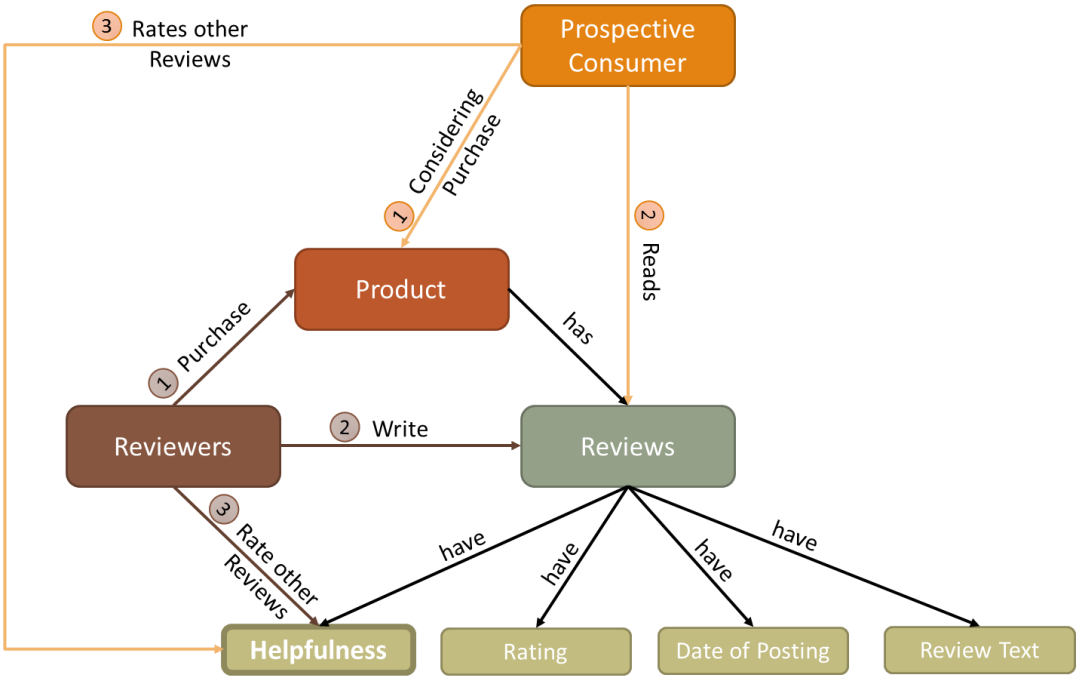


Figure 1. An overview of online review platforms (ORPs) and the associated work flow

According to a recent US nationwide surveys, writing, usage and trust in online consumer generated reviews have been steadily growing and consumer reviews posted online are currently the third-most trusted form of advertising [3]. Online review platforms (ORPs) have become an important information source that influences purchase decisions amongst prospective consumers [4–6]. Many studies have reported that consumers are often motivated to access product review information to make well informed purchase decisions [7, 8]. Research has also

shown that many prospective consumers use online product reviews[1] as information sources and those reviews influence product sale [9].

1.1 Problem with Online Consumer Reviews

The number of online consumer reviews generated on ORPs has been increasing at a rapid pace [9], especially for popular products and services, resulting in information overload for prospective consumers [10, 11]. The cost of information search from reviews may outweigh the benefits gained from reading numerous reviews, which decreases the effectiveness of online reviews as decision aids [12] for prospective consumers. It necessitates the need of identifying the relevant reviews from a large amount of reviews available [13]. Relevant reviews that can be helpful to prospective consumers are usually buried in the huge number of reviews [14], making it difficult and time consuming for the consumers to wade through and examine those reviews in order to find helpful information.

To overcome this problem, some online review platforms, such as Amazon.com, Bestbuy.com, and Yelp.com, provide some aggregated, summarized, or categorized views of online reviews, such as categorizing reviews based on (1) the helpfulness of reviews voted by fellow consumers, (2) the star rating of products, and (3) the date that reviews were posted (see Figure 2). Given the large number of reviews, however, such global review categorization and summarization methods are limited and inadequate for prospective consumers to identify relevant and helpful reviews. For example, the helpfulness votes of product reviews are sparse and skewed in nature.

[1] The terms "product review" and "consumer review" have been interchangeably used in this proposal

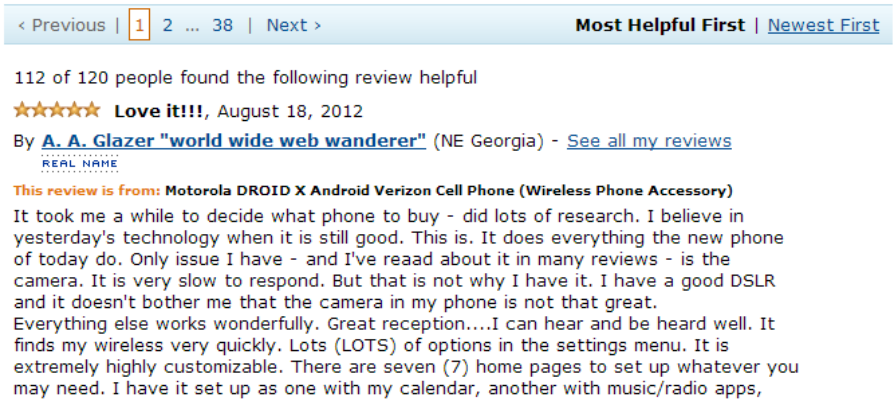


Figure 2. Example of review ranking provided by Amazon.com: (i) Most Helpful First and (ii) Newest First

Consumers may tend to only read top reviews that have received many helpfulness votes and in that process those top reviews keep garnering more helpfulness votes, which leads to the rich-get-richer effect. Thus, a lot of other reviews having relevant and helpful information about a product but have received fewer helpfulness votes may never get the opportunity to be read.

Researchers have conducted numerous studies to automatically determine the helpfulness of reviews by analyzing review content characteristics [10, 15–17]. This process helps identify many un-voted and/or less-voted reviews containing helpful information about a product. However, one of the major limitations of those studies is that helpfulness of individual reviews is predicted based on an assumption that people would perceive helpfulness of the same reviews equally, which is not always true. In many cases, individual consumers may have different preferences of or emphases on different product features when they make a purchase decision. When purchasing a mattress, some consumers may consider firmness as the most critical feature of a mattress, while others may prefer a mattress containing memory foam. Therefore, the

same reviews, depending on which product features were commented, could be helpful to some consumers but not necessarily to others. Similarly, consumers who prefer screen size, resolution, and battery life of a smartphone may mainly look for helpful reviews that particularly commented on those features of a target smartphone. As a result, it will be beneficial for prospective consumers if they can be presented with an ordered list of reviews that are ranked based on the relevance of review content to prospective consumers' preference on product features.

1.2 Personalization of Online Review Rankings

To address the above problem, it is essential to develop a personalized framework that would determine the helpfulness of individual review based on the product features of a product (i.e. personal preferences of the consumers). It will be beneficial for prospective consumers, if they can be presented with an ordered list of reviews that is ranked based on the relevance of review content to prospective consumers' preference on product features.

Some recent studies have explored personalizing online consumer reviews [13] [18]. However, they predicted review helpfulness based on consumer type (e.g., expert vs amateur), social activity of a consumer (e.g., as a reviewer and a rater), and a consumer's past product purchases rather than his/her product feature preferences. None of existing studies on review helpfulness determination has focused on predicting the helpfulness of reviews by taking consumer preferences into consideration. Accordingly, there has been little empirical research that examines the effectiveness of reviews presented in a personalized way based on their product

feature preferences. This dissertation research aims to fill this knowledge gap, by developing a personalizing review ranking framework to help consumers get more relevant and helpful reviews based on their product feature preferences and by providing a better understanding of the effectiveness of the review ranking personalizer through an empirical study.

1.3 Research Questions

Because not all reviews are equally relevant and helpful, extant research has suggested that it would be important to identify what reviews are influential to a consumer's purchase decision, and to personalize the presentation of those relevant reviews to meet the needs of a consumer [19, 20]. These two things lead us to the following major research question:

Whether organizing reviews in a systematic order that aligns with prospective consumers' specified product feature preferences could be more useful to prospective consumers than organizing reviews by their peer reviewed helpfulness votes that do not necessarily meet the consumers' specified needs?

More specifically the following two questions are examined in this dissertation research.

Question 1: What is an effective method for personalizing review ranking?

The development of an effective mechanism for creating online personalized review rankings is challenging. It needs various input factors (e.g., product reviews and review content characteristics, and preferred product features), advanced modeling

approaches, and robust design of a personalization framework. To answer this research question, a comprehensive framework that incorporates several technical approaches and related theories, is developed for creating a personalized review ranking system.

Question 2: Can the proposed review ranking personalization framework provide more helpful reviews to consumers based on their personal product feature preferences?

Practitioners and researchers need to better understand how online review ranking personalization will help consumers make better and quicker purchase decisions. In this research, an user study is performed that gather empirical evidence to demonstrate the effectiveness of personalizing review ranking systems by answering the following question: (i) how much relevant are the reviews to the consumers provided by the personalized review ranking system based on their product feature preferences, (ii) how much knowledge about a product do consumers gain by reading the reviews provided by the personalized review raking system and (iii) how much satisfied are the consumers with the information content of the reviews provided by the personalized review raking systems.

These questions are the driving force behind this dissertation research.

1.4 Research Objectives

In particular, the research aims to meet the following three objectives.

Objective 1: Automatically predicting helpfulness of online product reviews based on individual product feature(s) of a consumer's interest

The first objective of this research is to develop an advanced statistical model to predict the helpfulness of reviews of a given product. The statistical model uses the identified product features of a consumer's interest to predict the helpfulness of all reviews, especially, newly published yet un-voted reviews for that consumer. The statistical model also determines the contribution of individual product features towards the helpfulness of reviews.

Objective 2: Personalizing the order of online product reviews shown to consumers

The second objective of this research is to customize the order (i.e. personalize) in which the reviews are presented to prospective consumers based on their specified product features preferences. A prospective consumer specifies certain product features that are important to him/her (e.g. battery life in a tablet) when looking for a product and this information influences the order in which the reviews are presented to him/her.

Objective 3: Conducting an empirical study to evaluate the effectiveness of the proposed review ranking personalizer

The third and last objective of this research is to evaluate the effectiveness of the framework by conducting an online user study. The web based user study system aims to compare the performance of the proposed review ranking personalizer with the existing state-of-the-art review ranking system, reviews ranked by helpfulness.

This research posits that performance of the proposed review ranking personalizer system enables a consumer to access more relevant and helpful reviews to the consumers than the state of the art helpful ranking system by considering their product feature interests while building the statistical model.

1.5 Key Contributions of the Research

This research uniquely analyzes the content of every consumer review of a particular product in an ORP for personalization. A user's personal product feature preferences are captured and applied to process product reviews so that the significance of each review from the point of view of a prospective consumer could be predicted. This way of processing product features, consumer reviews, and capturing prospective consumer's preferences leads to review personalization.

The research makes several primary research contributions as discussed below.

Contribution 1: Helpfulness prediction of online product reviews based on consumers' dynamically changing product feature preferences

One major objective of the proposed research is the development of a helpfulness prediction model based on several features of a product that predicts and recomputes helpfulness of all reviews of that product for both already voted as well as un-voted reviews. It has been found that review helpfulness votes are very sparse in real review data sets [13, 18] and thus building a review helpfulness prediction model helps ORPs better present their reviews to consumers. There are some review helpfulness prediction models available in literature [15, 21, 22]. These studies have calculated

the helpfulness of individual reviews using either classification or regression-based prediction methods without considering a consumer's preference on specific product features of a product. These studies have also not considered the dynamic change of consumers' preferences for different product features while predicting the helpfulness of the reviews. One of the unique contributions of this research is the prediction of helpfulness for all individual reviews using a Latent Class Modeling (LCM) approach. More specifically, an advanced statistical model, latent class regression (LCR), is applied for helpfulness prediction of reviews that takes into account the dynamic change of consumer product feature preferences when providing personalized ranking of reviews. The calculation of helpfulness for all the reviews in turn help bringing relevant as well as better quality reviews to the forefront of the prospective consumers, based on their individual product feature preferences.

Contribution 2: Information overload reduction for prospective consumers by providing personalized review rankings

The research ranks product reviews based on the prospective consumer's specified product feature preferences by developing a *review ranking personalizer*. There have been very little research in the area of review personalization [13, 18]. However, those studies have not considered a consumer's product feature preferences while predicting the helpfulness of reviews and have not shown any ranked product reviews to consumers. Aligning the reviews with the preferences of a prospective consumer for particular product features helps him/her in dealing with information overload. The review ranking personalizer allows consumers in accessing more relevant

reviews in a time-effective manner by not having to search through all reviews. The proposed system also provides self-control to consumers in selecting and viewing reviews based on their specified product feature preferences. Enabling the consumers to view reviews based on their specified preferences, can make the prospective consumers comprehend reviews more carefully, thereby take the right and smart decision while making online purchases. The review ranking personalizer does not need social or historic data from consumers to personalize the review ranking, but only depends on the stated product feature preferences to present the reviews in a personalized order.

Contribution 3: Performance evaluation of the developed personalized review ranking framework by conducting an end user evaluation

Most of the studies in the helpfulness prediction domain have considered a model based approaches to evaluate the quality of the reviews predicted by the helpfulness prediction models [18, 23]. However, such approaches never consider user's direct feedback on the quality of the reviews predicted for him/her. On the other hand, user centric system evaluation can capture the direct attitude of the users towards the helpfulness of the reviews predicted by helpfulness prediction models. Based on this notion, this research has taken a user centric approach to evaluate the performance of the proposed review ranking personalizer (i.e. FDPPR framework). The system is evaluated based on the feedbacks (about the quality of the reviews on several aspects) from many end users. By conducting a user study, this research helps in evaluating

the efficacy of the proposed framework that can be adopted by ORPs to provide personalized review rankings to the consumers.

The remainder of this dissertation is organized as follows. In Chapter 2, a comprehensive literature review on online review analysis, latent class modeling and system evaluation is presented. Chapter 3, discusses about the feature driven personalization of online review (i.e. FDPPR framework) system architecture in great details. Chapter 4 presents a case study based on FDPPR framework. Chapter 5, presents the performance evaluation of FDPPR that includes hypotheses development, and user study design. In Chapter 6, the results of the user study are presented. Finally, Chapter 7 concludes this research by discussing implications of the research, limitations, and future research directions.

Chapter 2. Literature Review

In this chapter, literature review in the area of consumer review analysis, latent class regression models, and evaluation of helpfulness prediction models is presented in detail. Section 2.1 discusses online review analysis, which includes a background literature review in the area of review helpfulness prediction, review personalization, review sentiment analysis and product feature extraction using LDA. Section 2.2 covers latent class regression model used in this research for review helpfulness determination and review personalization. Section 2.3 provides a brief literature review about performance evaluation of helpfulness prediction models and the user-based studies performed on search engine ranking systems. Section 2.3 also presents a brief literature review on analytical hierarchical processing and kendall tau distance measure that are used in this research to evaluate the proposed review personalization framework. Section 2.4 summarizes the limitations of the current review helpfulness prediction/personalization approaches.

2.1 Online Consumer Review Analysis

The research on online consumer reviews has been steadily increasing in the past decade with a primary focus on the impact of online reviews on consumer's behavior and perceptions such as consumer purchase intention, satisfaction, and revisit intention [24, 25], the dynamic relations between online reviews and sales/revenues of e-marketplace [26–28], and the motivation and methods of online review communication and transmission [29, 30]. For example, in the domain of consumer purchasing intention, numerous studies analyzed the effect of on online reviews

quality (valence) and quantity (volume) on customers purchase intention [31–33]. Studies on sales and revenues domain examined the effect of consumer review ratings on the sales revenue of the products/services and found a substantial relationship between them. For example, one study [34] showed a positive effect of online book reviews on book sales using data on Amazon.com. In a similar study by Liu [35], a significant relationship was also shown between online movie reviews ratings and the box office revenues. In the rest of this section, prior related work on review helpfulness prediction and personalization is presented.

2.1.1 Online Consumer Review Helpfulness Prediction

More recently, research on automatically assessing and predicting helpfulness of online consumer reviews is gaining a lot of attention from the research communities. Helpfulness of an online product review reveals how consumers perceive the relevancy and value of a review. In most of the studies, review helpfulness has been calculated using various textual (e.g., length of a review, number of adjectives/nouns used, subjectivity of reviews) and/or social (number of past reviews, reputation of an author, past average rating by an author) factors extracted from the review textual content and review metadata. Some of the commonly used predictive features of helpfulness are:

- Review length: Review length (or depth of the review) has been considered as one of the important review related factors strongly associated with the perceived usefulness of the reviews by many previous studies. Most of the researchers [17, 36–39] have found a positive association between the length

of a review and the perceived usefulness of a review. Some research has indicated an existence of a threshold effect of length as well.

- Age of the review: Another important review factor that has contributed in determining the helpfulness of the reviews is the age of the reviews (i.e. days elapsed after the review being posted). Most researchers have concluded that age has a positive impact on the helpfulness [15, 39–41].

- Review ratings: Review ratings (usually from 1 star to 5 star) have shown a mixed response while studying the effect of review ratings on the helpfulness of the reviews. While some of the studies [39, 42, 43] have shown a positive effect of the review rating on the helpfulness of a review, a negative effect has also been revealed by many studies [40, 44].

- Sentiments: Cao, Duan, and Gan (2011) have analyzed and compared the effect of semantic characteristics and the other characteristics of the reviews on the helpfulness vote counts [11]. The research has concluded that reviews having extreme opinions get more helpfulness votes than the reviews having mixed or neutral opinions. Li et.al [45] have drawn similar conclusions as well. Salehan and Kim have investigated the effect of review sentiments on the helpfulness of the reviews [37]. The finding of the study has shown that reviews that contain more positive sentiment in the title get more readerships and reviews having neutral polarity in the text are perceived to be more helpful. The length and longevity of a review increases the readership and helpfulness of a review.

- Emotions: Emotions have been shown to have a role in determining helpfulness as well. Researchers in [46–49] examine the relationship between emotional review content (angry, anxiety) and helpfulness ratings of reviews and conclude that reviews with more content indicative of anxiety are helpful than the reviews containing content indicative of anger.

- Review extremity/readability/expression density and diversity : Review extremity has also been studied by Mudamb & Schuff in [24]. The study found that review extremity affects the perceived helpfulness of the reviews and the effect of review extremity on the helpfulness of the review is moderated by the product type. Similarly review readability has also been studied by [50, 51]. These studies have shown a strong impact of review readability on the helpfulness of the reviews. In another study [52], Willemsen et.al have studied the effect of review argument density and diversity on perceived helpfulness of the reviews and have found a positive impact.

- Information disclosure: Studies on information disclosure of a reviewer (identity-descriptive information displayed on review platforms such as real names, summaries of past contributions, self-photo, location, reviewer identity) have shown to have an impact on the helpfulness of reviews [6, 36, 50].

- Reviewer's expertise level: Reviewer expertise, measured by total number of reviews by the reviewer on a topic has shown a positive effect on the helpfulness of reviews [53]. Similarly, previous research focusing on the

effect of reviewer's expert level (e.g. ranks and special badges such as top- 50 reviewer or top-100 reviewer etc.) on the helpfulness of reviews have found a positive association between the two.

- Reviewer's social network: The reviewer's social network has also been studied to measure the effect on the helpfulness of reviews. Prior studies such as [36, 40] have found a positive relationship between reviewer in-degree/out-degree centrality, reviewer's fan, reviewer's follow numbers and the helpfulness of reviews.

Various classification and regression models have been used to develop models for helpfulness prediction. Classification based methods have been used to classify reviews into different classes. For instance, reviews can be classified into multiple classes such as high, low, or medium quality [21] or helpful or not helpful classes [54] based on those textual characteristics of reviews and posting behavioral characteristics by using classification algorithms such as Support Vector Machine (SVM). When using regression based prediction models, regression techniques such as Support Vector Regression (SVR) [22] and linear regression (LR) [15] have been widely used to predict the helpfulness of reviews. In [22] , SVR was used to evaluate the quality of a review by considering structural (e.g. html tags, review length), semantic (e.g. product feature mentions), syntactic (e.g. percentage of verbs or nouns in a review), metadata (e.g. star rating) and lexical features (e.g. n-gram) of a review. Ghose and Ipeirotis [15] applied linear regression (LR) with a variety of textual (review subjectivity, number of sentences in a review, number of words in a review)

and posting behavioral features (e.g. rank of the reviewer, number of past reviews) to predict the helpfulness of a review.

2.1.2 Online Consumer Review Personalization

So far only a few studies have explored methods for presenting reviews to consumers in a personalized way. For example, one study developed a series of probabilistic graphical models based on Matrix Factorization and Tensor Factorization to predict personalized reviews [18]. The study used details of the reviewers (e.g. past review ratings, number of reviews previously posted etc.), characteristics products, and extracted unlabeled latent features from the text of reviews to predict review helpfulness ratings. A similar study [13] also developed a method for predicting helpfulness ratings of reviews for individual consumers by exploiting context awareness for inferring unknown helpfulness ratings automatically. Four types of context, i.e., author context (i.e. reviewer who writes a review), rater context (i.e. rater who rates a reviewer), connection context, and preference context, were then mathematically formulated. A context-aware helpfulness prediction framework (CAP) was developed using a factorization model based on content context and various types of social context of the reviews to predict the helpfulness rating of each review. However, the major drawbacks of these studies are that they have not considered prospective consumer's product feature preferences while recommending helpful reviews (i.e. predicting helpfulness ratings of reviews in a personalized way) to consumers. The review recommendation has been entirely based on the consumer's past history with the ORP system (which includes past ratings, social connections, and other products reviewed). Research have shown that many prospective consumers

never rate the reviews [55]. This makes it difficult for these existing models to personalize reviews. Also, in those review personalization models, temporal effects (i.e. dynamic change of consumer preferences) have not taken into consideration while predicting helpfulness of reviews for each prospective consumer.

2.1.3 Sentiment Analysis of Online Consumer Reviews

Sentiment analysis is used to determine the sentiments (positive, negative or neutral emotions towards an entity) of a consumer expressed in a text data. There are two kinds of sentiment extraction processes available to extract sentiments from the text documents automatically, (1) classification based [56] and (2) lexicon-based [57]. In classification-based approaches (statistical/machine learning), a supervised classifier is built from labeled instances of texts or sentences. Most of the classification based sentiment analyses use Support Vector Machine classifiers that are trained on a domain specific data set using different features such as unigrams or bigrams, and with or without part-of-speech labels [56, 58–60]. Although such classifiers achieve high degree of accuracy in predicting the polarity of text documents in a particular domain on which they are trained on, they perform very poorly in other domains. Also, these classifiers do not take the effect of linguistic context such as negation (e.g., not good) and intensification (e.g. very good) into consideration while determining the polarity of the text documents [61]. In contrast, lexicon-based approaches determine the sentiment of a document by aggregating the sentiments of words/word phrases in the document. Lexicon based methods are very simple and seem to address the issues with classification-based methods by creating sematic rich dictionaries of words. There are many dictionaries available for sentiment detection

such as WordNet-Affect[2], SentiWordNet[3], and Opinion Finder [62]. Each dictionary includes multiple words and their associated sentiment scores that represent sentiment strength or just positive/negative polarity. For example, Hu & Liu [63] identified polarity for adjective words of a review text using WordNet.

Sentiment analysis has also been applied to determine polarity of text documents at different level of granularities that includes document level, sentence level, and aspect level. In document level sentiment classification, a sentiment classification algorithm classifies a whole text document as either positive or negative [56, 57]. For instance, for a product review, sentiment analysis finds out the overall positive or negative sentiment of the review about a product. In a sentence level sentiment classification [58, 59], each sentence of a text document is classified as positive, negative or neutral. In an aspect level sentiment classification, a fine-grained analysis is conducted to find out the sentiments towards different features or aspects of entities (e.g. picture quality of a camera). Aspect level sentiment classification systems first discover the aspects of an entity and then determine whether the opinion on each aspect is positive, negative or neutral [63, 64] .

2.1.4 Product Feature Extraction from Online Consumer Reviews

Discovering product features from consumer reviews can provide more fine-grained information and provide meaningful insights by organizing and indexing reviews based on extracted product features commented in the reviews. Various studies have used different techniques to discover product features form consumer reviews. The

[2] http://wndomains.fbk.eu/wnaffect.html
[3] http://sentiwordnet.isti.cnr.it/

product features have been usually extracted using one of the following techniques (1) string extraction-based methods, and (2) topic modeling.

Initially, string extraction-based methods were developed to extract features (aspects) from a text corpus. Those methods identified frequently appearing noun phrases in a text document as features. For example, Hu and Liu [63] proposed an association mining algorithm to identify features in product reviews. First, a linguistic parser parsed each review to split text into sentences. Then Part-of-Speech (POS) Tagging was performed to identify if a word is a noun, verb, adjective, etc. Next, association mining was carried out to find frequent "features (set of words or a phrase (here noun phrases identified by POS)) that occurs together in some sentences)" using CBA Miner. Wang et al. [65] proposed a bootstrapping based algorithm (aspect segmentation algorithm) to discover aspects (features) from review text. First, the algorithm split a review document into sentences and determined the initial aspect annotation by assigning each sentence to an aspect that had the maximum term overlapping with the sentence. Next, aspect dependency was calculated by Chi-square statistics and high dependent words were put under the corresponding aspect.

Those methods, however, have inherent limitations of putting several constraints such as compactness pruning and redundancy pruning [63] on noun phrases with high occurrence frequencies to identify product aspects. By exercising such constraints in the text analysis process, those algorithms not only become complex, but also identifies many non-aspects mistakenly. Those methods also miss the low-frequency aspects as well as their variations. Further, those statistical approaches require manual

tuning of various parameters, which makes it difficult to apply those techniques to text documents in different domains. Another real challenge with string extraction-based methods is that the number of aspects increases as text size increases. This makes it difficult and inefficient to automate the whole process of applying those algorithms to large review corpora.

To address the problem of handling large-size documents for aspect extraction, many researchers explored topic modeling techniques to overcome the limitations of string extraction-based methods by automatically learning model parameters from review text. Topic model based methods aim at discovering latent aspects (or topics) that reside implicitly in text documents using probabilistic statistical models [27-29]. Topic models are a part of probabilistic modeling in natural language processing and machine learning. In generative probabilistic modeling, data are treated as arising from a generative process that includes hidden variables and calculates a joint probability distribution over both observed and hidden random variables.

The initial standard topic model is the Probabilistic Latent Semantic Analysis (PLSA)[66], a probabilistic variant of LSA. PLSI assumes that a text document is generated using a mixture of K topics. However, a PLSI model does not make any assumptions about how the weights of each topic are generated, making it difficult to test the generalizability of the model for new documents. In addition, a model used in a probabilistic latent semantic analysis could run into severe over-fitting problems as the number of parameters grows linearly with the size of a text corpus [69]. Latent Dirichlet Allocation (LDA) [67] has been one of the most commonly used topic

modeling method. LDA is similar to PLSA, except that in the former, the topic distribution is assumed to have a Dirichlet prior [67], resulting in more reasonable mixtures of topics in a document. The basic idea behind LDA is that it posits that each document is a mixture of a small number of topics and that each word's creation is attributable to one of the document's topics.

LDA has been applied to consumer reviews successfully to find latent aspects from reviews [9], [34-37]. One of the major advantages of LDA is its modularity which allows researchers to extend it for finding complex structures in text documents (e.g. finding topic correlation, finding topic hierarchy etc.). One of the extensions of LDA is the Correlated Topic Model (CTM) [74] that assumes presence of one latent topic may be correlated with the presence of another topic. The topic proportions in CTM are drawn from a logistic normal distribution rather than a dirichlet distribution. Another extension of LDA is the dynamic topic model [75] which takes the ordering of documents in account and gives a richer posterior topical structure in the text than LDA. A family of probabilistic time series models is developed to analyze the evolution of topics in large document collections. One more extension of LDA is pachinko allocation machine (PAM) [76] that allows the occurrence of topics to exhibit correlation (e.g. a document about geology is more likely to also be about chemistry then it is to be about sports). In PAM, the concept of topics is extended to be distributions not only over words, but also over other topics. Titov & McDonald [69] extended the basic LDA model by assuming and representing a document as a sliding window, each covering a number of adjacent sentences within a document and proposed a multigrain topic model that models and find local and global topics in

text documents. In LDA, inference is performed to learn topics distribution for each document, associated word probability distribution, and the particular topic mixture of each document. Originally, the variational EM (Expectation Maximization) [67] approach was proposed by David et.al. Later various approximations have been considered [67, 77, 78]. Another approach is to use a Markov chain Monte Carlo algorithm (MCMC) for inference with LDA [79]. Gibbs Sampling [78], is an example of MCMC for the posterior inference of the parameters (α, β). The values of α and β are considered to be 50/(number of topics(k)) and 0.1, respectively, as suggested by [78].

2.2 Latent Class Model and Latent Class Regression (LCR)

Social and behavioral sciences usually involve many constructs that are not fully observable or measurable [80]. For example, depression and self-esteem are constructs that investigators widely agree to exist, yet are not directly measurable. Symptom checklists or surveys consisting of multiple items are often used to study such constructs to capture the underlying latent construct. However, the inherent unobservable nature of a latent variable makes it difficult to correctly measure them using such methods. That in turn leads to the measurement error (the discordance between the underlying true construct and the observed indicators). The latent variable modeling was developed in order to appropriately account for the measurement error arising from the study of latent variables. One widely used latent variable model is latent class model (LCM) [81–83]. LCM models a categorical latent variable based on multiple, observed indicators U_1, U_2, …, U_j, such that each

individual entity belongs to exactly one of C latent classes, denoted C_1, C_2, …, C_C. The latent classes are defined by response patterns among the indicator items. For example, one class may have a low propensity for endorsing all items while another class is defined by high propensities for endorsing all items. LCM is used to explain the interdependency of the categorical observed variables by introducing the explanatory latent variable. Initially LCM was proposed through a survey to study social attitudes of individuals in which typologies were built using sets of dichotomous observed variables [82]. Then it was extended by Goodman [83] who developed maximum likelihood estimation procedures that could deal with polytomous variables.

A latent class regression (LCR) model generalizes LCM by allowing for auxiliary variables (covariates) to be related to latent classes [84]. In a LCR model, apart from the variables used to identify latent classes, covariates can be included in the model affecting the class membership [55, 56]. LCR typically assumes that covariates do not have any direct effect on the indicators Ui,…. Uj, but rather is only associated with the indicators via class membership. LCR models are used to group entities of interest into different classes where the entities within each class have similar values on a set of observed indicator variables. LCR starts with the creation of individual linear regression equations for all classes. LCR differs from the classic linear regression as it identifies multiple alternative equations that can be applied to different latent classes. These multiple regression equations explain the variance in the data better than a single regression equation. LCR also calculates the probability

that individuals belong to each class and a regression equation for each latent class is developed based on weighted probabilities.

LCR models have been used in many areas such as marketing [87], sociology [88], psychometrics [89], medical research [42, 43], and psychosocial [44, 45] etc. Recently, LCR models get increasing attention from various research communities, as those models can yield powerful improvements over traditional approaches to clustering, and/or regression/segmentation [94–97] as well as multivariable biplots and related graphical displays. The major drawback of traditional approaches such as discriminant, simple regression and log-linear analysis is that they describe only relationships among observed variables. LCR models are different from these traditional models as they contain parameters that can relate one or more observed variables to discrete unobserved (latent) variables in the data. Also traditional models rely on various modeling assumptions such as linear relationship, normal distribution, homogeneity, which are often violated in practice and thus introduce biases in the models [98]. In contrast, LCR models are not restricted by those model assumptions, thereby reduce biases associated with data. In addition, LCR models can handle various types of variables, such as nominal, ordinal, continuous and/or count variables, in the same analysis. In LCR, latent classes and external variables (covariates) are assessed simultaneously with the identification of the clusters that leads to improve cluster description. This eliminates the use of any second stage of analysis such as discriminant analysis after the first stage of analysis (traditional clustering) to describe the clusters.

2.3 Evaluation of Helpfulness Prediction Models

System evaluation is an important part of information system research that tests the performance of the developed system. It validates the output of a system and determines the feasibility and effectiveness of the system. In previous studies, model-based approaches have been adopted for evaluating the performance of helpfulness prediction models. Most of the studies used K fold cross validation techniques to determine the quality of the helpfulness prediction. For instance, in the study by Chen and Tseng [22], helpfulness ratings were ranked according to the time points when they were published in a chronological order and then whole data set was equally split into two parts. 50% of them were considered as the training dataset and 50% of them were treated as the testing dataset. The helpfulness prediction model trained on the 50% data and tested on the rest 50% test data. Root Mean Squared Error (RSME) was then used to evaluate performance of the system. Similarly, Moghaddam et al. [18] used 10-fold cross validation on the dataset to measure the performance of their system. In each fold, 90% of the data were used as the training dataset and the remaining 10% as test dataset. Root Mean Squared Error (RMSE) was used as the evaluation metric. In another study [99], helpfulness prediction accuracy of the developed model (back-propagation multilayer perceptron neural network) was examined by comparing it with the multivariate regression analysis. V-fold cross validation approach was used on the dataset and the results from both models were compared. The whole dataset was split into a number of subsets where each subset was treated as a test sample one by one while leftovers were treated as the training dataset. 12 neural network models and regression models were trained or estimated.

Mean Square Error was aggregated for the 12 test samples for both the models and the results were compared. Kim et al. [22] used 10 fold cross validation to evaluate their developed review ranking system using Support Vector Machine (SVM). SVM was trained by 9-fold, while in the remaining test fold each product's reviews were ranked based on the SVM prediction. Spearman's correlation coefficient was used in the study to correlate the ranking of the reviews with the gold standard ranking (user helpfulness votes on Amazon.com). Similarly, O'Mahony and Smyth [55] applied 10 fold cross validation on three classifiers including their (classifies reviews as helpful or non-helpful) to measure and compared the performances of three systems using receiver operating characteristic (ROC).

Although model based performance evaluations are conducted to test the performance of the constructed models, such approaches fail to assess real quality of the result generated by these models. In contrast to model based evaluation, user based evaluation approaches (usability studies) capture direct feedback about the quality of the results from end users. This research develops a personalized helpfulness prediction model that has focused on providing personalized review ranking systems to the consumers based on their product feature preferences. So far, there are no studies in the literature that have conducted user-based evaluation of the helpfulness prediction models. A related research area close to this research is the search engine ranking system that ranks web documents based on their relevancy to the user's queries. Several user studies have been conducted in this domain to evaluate the ranking of search results using human relevance judgement. For example, in one study by Su et al. [100], users were asked to choose and rank the five most relevant

items from the first twenty results retrieved for their queries. Similarly Hawking et al. [101] and Chowdhury and Soboroff [23] evaluated the effectiveness of several web search engines using reciprocal rank of the relevant document, a measure closely related to ranking. Vaughun [102] compared human rankings of 24 participants with those of three large commercial search engines, Google, AltaVista, and Teoma, on four search topics. Beg [103] compared the rankings of seven search engines on fifteen queries with a weighted measure of the users' behavior based on the order the documents were visited, the time spent viewing them and whether they printed out the document or not.

User centric studies provide an avenue for high-quality objective and subjective evaluation of the system [104]. Due to this reason, in this research, a user centric approach is considered to evaluate the proposed personalizing review ranking framework. The quality of result (review) ranking is calculated based on a continuous relevance ranking (from most relevant to least relevant) by human subjects (which is captured by pairwise comparing a set of reviews by Analytical Hierarchical Processing (AHP) [105]). The effectiveness of the system is measured by capturing the difference in the quality of result ranking by different ranking systems (system generated vs user generated) using Kendall Tau Distance [106]. Kendall Tau Distance is a metric that counts the number of pairwise disagreements between two ranking lists. The lower the distance, the closer a ranking system is to the human ranking and thus the better the ranking performance. The next two sections present a brief introduction of AHP and Kendall´s Tau Distance.

2.3.1 Review Ranking Using AHP

Analytical hierarchical processing (AHP) is a popular multi-criteria decision-making method. It was originally proposed by Saaty [107]. It has been used as an effective tool for dealing with complex decision-making process and helps decision makers to make the best decision by setting priorities. In AHP, pertinent data are derived by using a set of pairwise comparison. The pairwise comparison method is used to measure the weight of two criteria by measuring the relative importance of one criteria over the other. By using a set of pairwise comparisons, AHP helps to capture both subjective and objective aspects of a decision. The pairwise comparison method was originally introduced in psychological research [108]. It was then improved mathematically by Saaty and served as the basis of the AHP [107, 109]. In each pairwise comparison, AHP uses simple linguistic phrases to get the qualitative data and uses a numerical scale to quantify the data. The scale indicates how many times more important one element is over another element with respect to a criterion. There have been several different numerical scales such as the Saaty scale [109], the geometrical scale [110] and the Salo-Ha¨ma¨la¨inen scale [111]. The Saaty scale has been the most commonly used scale in many applications. In addition to that, AHP incorporates a useful technique called consistency ratio (CR) for checking the consistency of decision maker's evaluations, thus reducing the bias in the decision-making process. AHP has been applied to wide verity of domains such as marketing [112, 113], energy [114, 115], medical and health care decision making [116, 117], research and development (R&D) project selection and resource allocation [118], etc.

2.3.2 Ranking Systems Comparison using Kendall Tau Distance

Determining which ranking is the best among a list of potential options is a complex problem, described by [119] as a Holy Grail search task. Kendall tau distance [120] has been widely used in many research comparing different ranking systems. Kendall tau distance (*d*) counts the number of pairwise disagreements between two ranking lists. The larger the distance, the more dissimilar are the two ranking lists. Kendall tau distance is ideally suited to situations where measurements are subjective or difficult in practice. This research uses Kendall Tau distance (*d*) to find the distance between the preferred ranking of the user and two system generated ranking systems.

2.4 Limitations of the Current Review Helpfulness Prediction/Personalization Approaches

As evident from the current research on helpfulness prediction and personalization of online consumer reviews, there has been some effort to provide relevant reviews to prospective consumers. However, the existing ORPS are still far from providing a personalized review ranking system to prospective consumers to help them find the relevant reviews based on their product feature preferences. First, there have been many approaches for automatically determining the helpfulness of reviews. However, there has been no attempt to develop a systematic approach that takes a consumer's interest for specific product feature of products while modeling and predicting the helpfulness of reviews. There are a few review personalization approaches that have focused on providing reviews to consumers in a personalized way. But they have used social and historical/ behavioral information of the consumers (armature vs expert,

raters vs authors, past purchases etc.) to predict helpfulness of the reviews rather than product features. Also, there have been a lot of helpfulness prediction models developed on online consumer reviews and empirically tested using model based evaluation approaches. However, there have been no empirical user studies yet that explain the effect of review helpfulness in providing relevant and helpful reviews to consumers.

Chapter 3. A Framework for Product Feature Driven Personalization of Online Product Reviews (FDPPR)

This chapter presents a generic and comprehensive theoretical framework for developing a "*review ranking personalizer*" that ranks online product reviews based on the product feature preferences of a consumer. The framework for "Product Feature Driven Personalization of Online Consumer Reviews (FDPPR)" is proposed to help researchers and practitioners develop a personalized review ranking driven by the product feature preferences of individual consumers. When a prospective consumer is looking for a product for a potential purchase, he/she usually reads a lot of online reviews about that product in advance. Prospective consumers usually have preferences for certain product features of a product. They access the suitability of a product for purchase based on reading reviews where these product features are commented. Consumers are currently faced with thousands of reviews for a single product to get the summary of a single product feature that interests them.

The proposed FDPPR framework provides a "*review ranking personalizer*" based on consumers' personal product feature preferences to show relevant reviews that adequately address a consumer's specific needs. In particular, FDPPR uses a LCR model that takes into account the dynamic change of consumer's product feature preferences when providing personalized rank of reviews to him/her. A prospective consumer is allowed to specify product features of his/her interest of a product and the "*review ranking personalizer*" ranks reviews to better suit the preferences of the consumer. This research also allows consumers to change their product feature

preferences at any time and uses the proposed review ranking personalizer rank reviews in different order so that consumers get a better understanding about the product. The FDPPR provides a personalized review ranking system to tailor the review ranking for a prospective consumer based on his/her product feature preferences.

The review personalization developed as part of this research is going to be beneficial for all prospective consumers, including the unregistered ORP consumers, as it personalizes review ranking without depending on consumers' past purchase history or past interaction with an ORP system. FDPPR provides a personalized approach to online product review analysis and presentation so as to better support consumers' purchase decisions.

This chapter is organized as follows, Section 3.1 presents the proposed framework and discusses the two primary aspects of the framework, namely offline review analysis and online review personalization. Section 3.2 presents the development of the offline review analysis module using a LCR model. Section 3.3 presents the online review personalization process.

3.1 Feature Driven Personalization of Product Reviews (FDPPR) Research Framework

The design, development, and evaluation of a user-centered personalized product review analysis and presentation will better support consumers' purchase decisions by presenting them with the most relevant reviews first based on their product feature

preferences. Figure 3 presents a generic and comprehensive theoretical framework called FDPPR for creating review ranking personalizer. The FDPPR framework provides a novel approach to building a review ranking personalization system using state of the art text analytics and the LCR model.

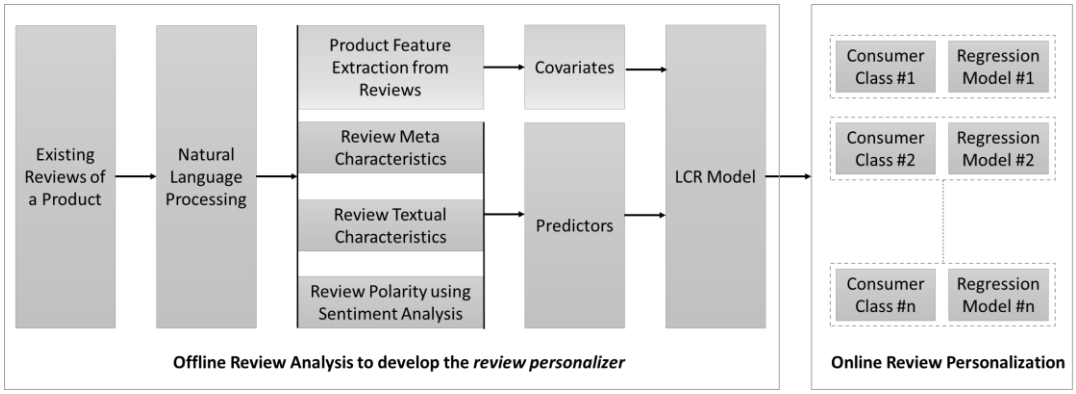


Figure 3. The Product Feature Driven Personalization of Online Consumer Reviews (FDPPR) Framework

There are four objectives of the proposed research (see Section 1.3) that include (1) characterization of consumer reviews in terms of product features, review stylistic characteristics, and sentiments, (2) predicting the helpfulness of online reviews, (3) personalization of online review order using the *review ranking personalizer*, and (4) a user centric evaluation of FDPPR.

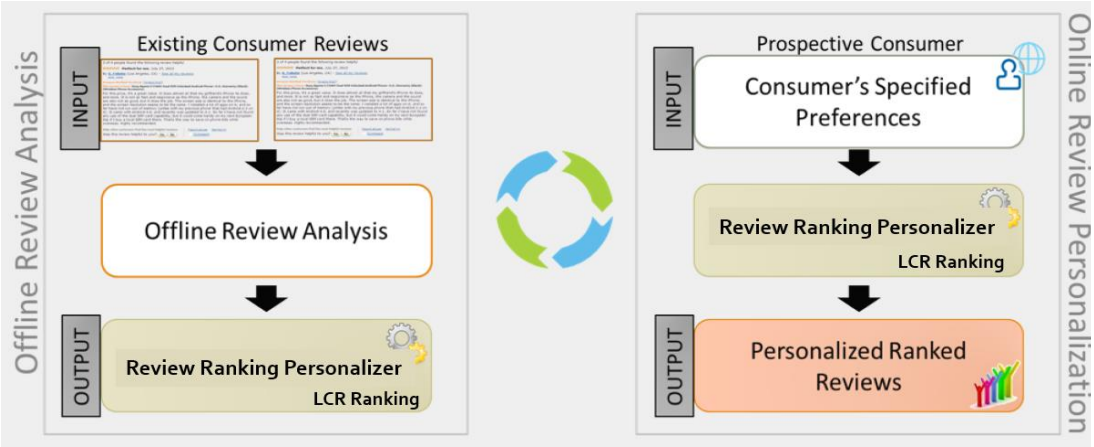


Figure 4. An Overview of the offline review analysis and online review personalization modules of FDPPR

To fulfill the first three research objectives, two separate but complementary review analysis modules are developed, namely the *offline review analysis* and the *online review personalization* (see Figure 4).

In the *offline review analysis* module, product features, stylistic characteristics of the reviews, and the sentiment associated with reviews are extracted from the review text using NLP. Once reviews are represented by the extracted features, a LCR model is developed (i.e. *review ranking personalizer*) and then used for online review personalization. In the *online review personalization* module, the developed *review ranking personalizer* is used to provide personalized ranked reviews to prospective consumers based on their specified preferences. The offline review analysis is run periodically to update the *review ranking personalizer* so that a prospective consumer will always get personalized order of reviews. Figure 5 shows the schematic of the review ranking personalizer that is used to develop personalized review ordering to adequately satisfy a consumer's product feature preference(s).

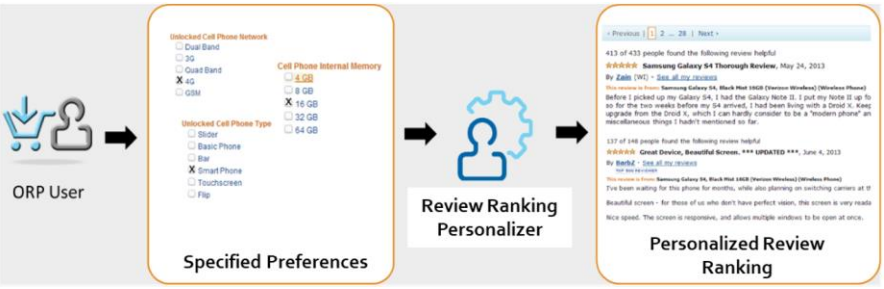


Figure 5. The review ranking personalizer for re-ordering reviews based on prospective consumer's specified preferences

A consumer specifies the product features that he/she is interested in and the review ranking personalizer uses that information and calculates a relevance score for every review of that product based on the LCR model. The calculated scores of the reviews are then used to sort the reviews in a descending order and displayed to the consumer. If the consumer changes his/her product feature preference, the same LCR model calculates a different score for each review based on the changed product feature preference of the consumer. This leads to a different order of reviews to suit the consumer's need. The ordered reviews help a consumer find and read the most helpful and relevant reviews quickly.

Table 2. Mapping between the offline review analysis and online review personalization module

Personalization Module	**Steps**
Offline Review Analysis	1. Product Feature Identification and Extraction from Review Corpus 2. Review Meta-Characteristics Identification and Extraction 3. Review Textual Characteristics Identification and Extraction 4. Review Polarity Determination (using Sentiment Analysis) 5. Development of the Review Ranking Personalizer using LCR
Online Review Personalization	6. Identification of Prospective Consumer's Class based on Specified Preferences 7. Re-ranking Reviews based on Prospective Consumer's Class

The FDPPR framework provides a consistent and traceable way for modeling the characteristics of thousands of reviews for a product as well as the consumers using natural language processing (NLP) and the LCR model. The *offline review analysis* module mostly encompasses the data preprocessing and collection steps (steps 1 through 5) and the LCR model development. The *online review personalization* module encompasses steps 6 and 7 (see Table 2).

3.2 Offline Review Analysis to Develop the Review Ranking Personalizer

To develop the offline review ranking personalizer using the LCR model, multiple review characteristics such as product features, stylistic characteristics, and sentiments are extracted from a set of reviews written about a single product using NLP. More importantly, in this research, several review stylistic characteristics that have been consistently used in previous studies on review helpfulness prediction (see section 2.1.1) are used in the development of the review ranking personalizer. The following sub-sections (see Section 3.2.3 to Section 3.2.4) discuss the steps followed to extract the review characteristics which are used to develop the LCR model. The last section (Section 3.2.5) discusses the details about developing the LCR model.

3.2.1 Review Meta Characteristics

Apart from the text in a review, every review has certain meta characteristics associated with it. The meta characteristics provide useful information about reviews (see Figure 6).

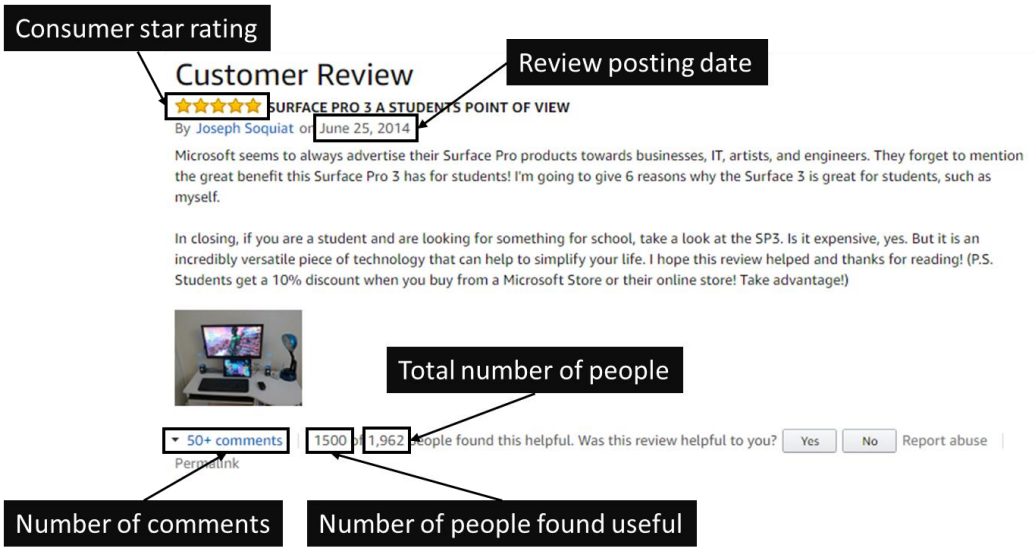


Figure 6. Anatomy of Review Meta-Characteristics

Some of the meta characteristics that are considered as part of FDPPR are:

1. *Review posting date*: Every review has a posting date associated with it that shows when the review was written. This value is converted into number of days since the review was written and was considered as part of the LCR model building.
2. *Consumer star rating*: Every review is usually accompanied by a consumer star rating. Most of the consumer ratings are usually in a scale from 1 to 5 where 1 is considered as lowest rating and 5 as highest rating.
3. *Number of people found useful*: Many reviews have an extra field that shows how many consumers found that review helpful. This is a very important field as the review has been vetted by other consumers. In this research, this variable is considered as a very important part of the LCR model.
4. *Total number of people*: This variable shows how many consumers read the review. A subset of the consumers who read the review mark it as useful.
5. *Number of comments*: This variable gives the number of comments posted for a given review.

Figure 7. Review textual characteristics present in a typical review from Amazon.com ORP

3.2.2 Review Textual Characteristics

The title and text of reviews (see Figure 7), which are part of a review corpus, are analyzed using natural language processing (NLP pipeline) to identify and extract textual characteristics from reviews.

The NLP pipeline in this research involves sentence splitting, tokenization, and stop word removal. As shown in Figure 8, the input to the system at this step is a set of consumer reviews and the output are the extracted textual characteristics. Processing reviews through a NLP pipeline helps take care of the variability usually found in review text for describing a product. For example, by performing stop word removal on the review text, the words like “the”, “is”, “at”, “which”, and “on” etc. are removed. This process greatly improves the search process to identify textual characteristics in the review corpus.

Figure 8 shows all the steps involved in parsing review text data and extracting textual features associated with the reviews. In the first step, a review text is split into individual sentences and then tokenized to get individual tokens (i.e., words or terms) from the sentences. Next, stop words are removed from the token set by matching each token with a list of stop words [121]. At the end, various textual characteristics are extracted from the cleaned review text data.

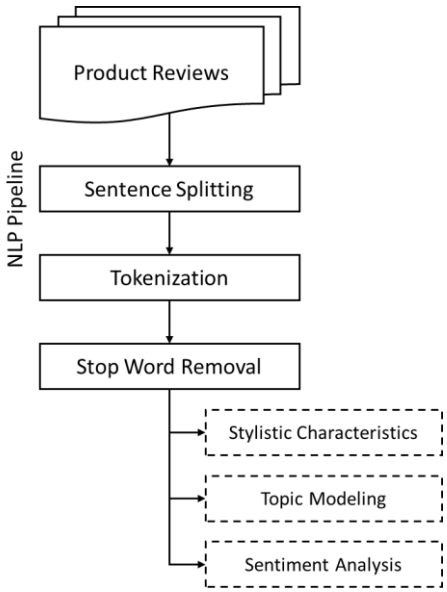


Figure 8. The NLP pipeline to preprocess the review text

3.2.3 Product Features Extraction from Reviews

In this research, a topic modeling-based method (LDA), is performed on the reviews to identify the product features that the consumers wrote about in their reviews. As discussed in section 2.1.4, LDA is useful in capturing and quantifying latent information presented in the text data in various domains and context. The simplicity and modularity of the basic LDA model have made it one of the most widely used topic modeling techniques. It has been applied to a variety of applications and also serves as building blocks in other powerful models. Thus, in this research basic LDA model is applied to identify and extract product features from consumer reviews.

Topic model based methods aim at discovering topics (i.e. product features) that reside implicitly in text documents (i.e. product reviews) using probabilistic statistical

models [27-29]. For each identified topic in LDA, a set of terms are identified that contribute (i.e. appear with a high probability) to the topic. In FDPPR, topic modeling is used to reduce the term space and identify product features that contribute towards the topics without starting with a fixed set of product feature taxonomy. These product features are subsequently used as covariates in the LCR model (see Section 3.2.5).

The LDA process is represented in Figure 9 using a standard graphical model notation. Each node in the graphical model denotes a random variable, while edges denote dependence between random variables. Shaded nodes in the graphical model denote observed random variables, and the un-shaded nodes denote hidden random variables. Rectangular boxes are "plate notations," which denotes replication. In a graphical model, while an outer rectangle box represents documents, an inner rectangle box represents the repeated choice of topics and words within a document. An inner solid circle represents an observable variable (each term in a document) in the LDA model. In a simple way, the graphical model can be interpreted as follows.

Step #1. For each document, a distribution over topics (θ_i) is randomly chosen from the dirichlet distribution (α)

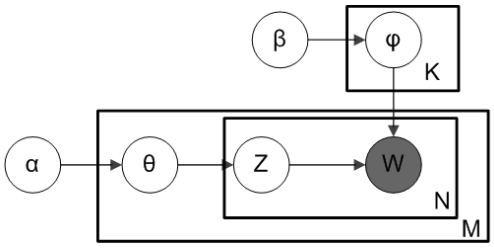
Step #2. For each topic, a distribution over vocabulary of words (φ_k) is randomly chosen from the dirichlet distribution (β)

Step #3. For each word W in the document i,

(i) a topic is randomly chosen from the distribution over topics(θ_i) in step #1

(ii) a word is randomly chosen from the corresponding topic distribution over the vocabulary (φ_k) in step #2

Graphical Model



Notations

M = number of documents
N = number of words in a document
α = parameter for per-document topic distribution
φ_k = word distribution for topic k
β= parameter for per-topic word distribution
θ_i= topic distribution for document i
z_{ij} = topic for the jth word in document i
w_{ij} = specific word (Observable variable)

Figure 9. A Graphical Model of LDA for text analysis

In this research, LDA analysis of the review corpus, created from real Amazon product reviews, is done using the Mallet library[4] (see Section 4.1.4). In the LDA analysis, individual product reviews are given as input and a pre-defined number of topics are extracted from reviews (see Section 4.1.4).

Term	Score	Term	Score
abandon	-2	aboard	1
abandoned	-2	absentee	-1
abandons	-2	absentees	-1
abducted	-2	absolve	2
abduction	-2	absolved	2
abductions	-2	absolves	2
abhor	-3	absolving	2
abhorred	-3	absorbed	1
abhorrent	-3	abuse	-3
abhors	-3	abused	-3
abilities	2	abuses	-3
ability	2	abusive	-3

Figure 10. Example of sentiment score in AFINN library

[4] http://mallet.cs.umass.edu/

Once the topics are extracted, the terms that contribute towards the topics are aggregated and a set of product features are extracted from the aggregated terms.

It should be noted that, one of the performance problems of LDA model is to determine the number of topics beforehand. To overcome this problem various mathematical metrics have been proposed to calculate the number of topics such as perplexity [67], empirical likelihood and marginal likelihood (harmonic mean method [76], annealed/mean field importance sampling [122], chib-style estimation [123], etc. These methods fit various models with different number of topics and the model having optimal number of topics is selected (i.e. in a data driven way). Although these methods have been applied to different text domains, previous research have also shown that, they are negatively correlated with the measures of topic quality [124]. Thus, many authors have suggested to take human inputs while determining or evaluating the performance of the topic models rather than optimizing likelihood-based measures [125, 126]. Also in topic modeling, topics are usually presented with manual post-hoc labelling for ease of interpretation in several research [127, 128]. In line with the previous research, this research has also performed a manual inspection of LDA model outputs with different number of topics to select the optimum number of topics and manual labeling to get the topic names from the highly probable keywords under each topic.

3.2.4 Review Polarity extraction using Sentiment Analysis

In this research, sentiment analysis is performed on each review at review level (i.e. document level) and the output sentiment scores are used as predictors in the

proposed LCR model of the FDPPR framework. Many previous studies have applied sentiment analysis to consumer reviews [17-20] to extract the sentiments expressed by consumers in those reviews. Research shows that, lexicon-based approaches are simple and overcome problems with classification-based methods (see Section 2.1.3). Thus, in this research, a lexicon-based method is adopted to determine the polarity of review documents (i.e. sentiment analysis) at the review level using the AFFIN[5] dictionary.

There are three steps involved in sentiment analysis, including opinion word extraction, polarity determination of opinion words, and calculation of the positive sentiment and negative sentiment scores associated with the review.

Step 1: Opinion word identification and extraction

In this step, words from the AFINN library present in review sentences are identified as the opinion words. Each review document is split into sentences, tokenized and stop words are removed. Then opinion words are extracted from the review sentences. This process continues for every sentence of a product review to extract all the opinion words.

Step 2: Determination of opinion words' polarity

After extracting opinion words from each review, the next step is to determine the polarity of each opinion word based on the word list available in AFFIN library (see Figure 10). AFFIN library contains a list of 1468 unique English words and phrases.

[5] http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010

Each word in AFINN has an associated sentiment score (integer between -5 (negative) to +5 (positive). Words that are not present in AFINN opinion lexicon list are discarded from the review as they may not be valid opinion words.

Step 3: Calculating the sentiment score

The positive opinion word count is treated as the positive sentiment score of the review, and the negative word count of the review is treated as the negative sentiment score of the review.

The output of sentiment analysis are added as predictor variables to the LCR model and contribute towards the helpfulness prediction. At the end of review polarity extraction, the results are stored with the individual reviews and are used in the development of the review ranking personalizer using the LCR model.

3.2.5 Building Review Ranking Personalizer using LCR

FDPPR uses LCR models for the development of the review ranking personalizer. As discussed earlier, LCR provides the following advantages over the other prevalent methods for *helpfulness prediction* such as classification and/or regression [15, 21, 22]. LCR uses a maximum likelihood-based technique to simultaneously identify the heterogeneity in the review types (cluster) and predict review helpfulness. The model fits are always better (or equal to) that those obtained using a two-step process, clustering followed by regression when the information provided to the model is equivalent [98]. Further, the maximum likelihood-based approach enables the collection of the statistical properties of the estimates (quality of the model fit, level of confidence a parameter value can have etc.). This is important as the relative

importance of the variables and the goodness of fit of the model are also obtained in the analysis.

Table 3. Variables that are collected for the building the review ranking personalizer

Type	Variable	Explanation
Product Features	Technical Features	Based on the product's technical features discussed in the reviews provided by the consumers. The number of variables will vary from product to product.
Review Characteristics	Star rating	The rating of a review that is given by the author of the review (i.e. the reviewer)
	People found helpful	The total number of people who found the review helpful
	Total number of people	The total number of people who read a review
	Days from first review	The number of days since the first review was posted
	Number of comments	The total number of comments for a review provided by other consumers
	Length (Number of Words)	The length of the review in terms of the number of words in a review
Review Sentiment	Positive Sentiment Score	The positive sentiment of a review based on the entire content of the review (i.e. review level positive sentiment)
	Negative Sentiment Score	The negative sentiment of a review based on the content of the entire review (i.e. review level negative sentiment)

This research incorporates the *specified preferences* of a prospective consumer to classify that consumer into a *consumer class* using the LCR model. A LCR model uses clustering and regression model simultaneously. Thus, there is a *regression model* associated with every identified *consumer class* identified by the LCR based clustering approach. Once a prospective consumer is identified to be part of a

consumer class, the associated *regression model* is used to calculate (i.e. predict) the helpfulness of all the reviews for that consumer. The reviews are then ranked by this calculated helpfulness score and presented to the consumer in a decreasing order. In other words, a unique personalized order of reviews is presented to each prospective consumer based on his/her product feature preferences.

In this research, a LCR model is developed to quantify the association between (i) product features, (ii) review stylistic characteristics, and (iii) sentiments of the review and the helpfulness of the review.

LCR works under the assumption that reviews cater to $\boldsymbol{C}$ classes of heterogeneous consumers; any two consumers who are in the same class are homogenous (i.e. have preference for the same product features and reviews), while any two consumers in different classes are heterogeneous. LCR then (1) identifies the relationship between a dependent variable (in this case, the number of readers who found the review helpful) and the potential independent variables (see Table 3) in each of the C classes, and (2) measures the probability that a prospective consumer would belong to a class C based on his/her specified product feature preferences.

Once the preferences of a prospective consumer are known, the consumer is shown a ranked list of reviews, that caters to the specific need of that consumer.

The workflow of LCR model can be expressed mathematically as follows:

Let $\boldsymbol{i} = \mathbf{1} \cdots \boldsymbol{N}$ be the index of reviews, $\boldsymbol{x} = \mathbf{1} \cdots \boldsymbol{X}$ be the index of variables extracted from consumer reviews (see Table 3), and $\boldsymbol{c} = \mathbf{1} \cdots \boldsymbol{C}$ be the index of latent classes,

- Let $\boldsymbol{Y_i}$ be the dependent variable that is a transformed measure of the helpfulness rating of review $\boldsymbol{i}$.

- Let $\boldsymbol{P_c}$ be the probability that a prospective consumer belongs to a latent class $\boldsymbol{c}$.

Then, the LCR model identifies the following relationships:

Regression Model $$\boldsymbol{Y_c = X\alpha_c + \varepsilon_c} \quad (1)$$

Logit Model $$\boldsymbol{log(P_c/(1 - P_c)) = f(X\gamma_c)} \quad (2)$$

Equation 1 represents the regression model, where $\boldsymbol{\alpha_c}$ is a $1 \times \boldsymbol{X}$ -dimension vector of coefficients for each variable of a consumer review for consumer class $\boldsymbol{c}$. Equation 2 represents the logit model where $\boldsymbol{\gamma_c}$ is a $1 \times \boldsymbol{K}$ −dimension vector of coefficients that enable the classification of consumers into $\boldsymbol{C}$ classes. Equation 1 and 2 together form the LCR model. Estimation of $\boldsymbol{\alpha_c}$ is performed using an expectation–maximization (EM) algorithm that identifies the helpfulness of any variables extracted from the consumer reviews for a prospective consumer.

The LCR model proceeds by assuming that a vector of observations Y_i arises from a population that is a mixture of $\boldsymbol{C}$ classes in proportions $\pi_1, \pi_2, \cdots \pi_c$ (where $c = 1 \cdots C$ is an index of classes). It is assumed that the membership of the observations in different classes is unknown a priori but $\sum_{c=1}^{C} \pi_c = 1$ and $0 \leq \pi_c \leq 1$; π_c can be interpreted as the point probability of a review belonging to class $\mathbf{c}$.

The LCR model then seeks to solve the following mixture latent regression model:

$$P(Y_i|X_{ik}) = \sum_{c=1}^{C} P(w = c|X_{ik})\, P(Y_i|w, X_{ik}) \tag{3}$$

Where $P(w = c|X_{ik}) = \pi_c$

and the probabilities $P(w = c|X_{ik})$ are parameterized and restricted by means of (logistic) regression models. In other words, the LCR model classifies consumers into consumer classes and develops regression models for each consumer class simultaneously.

The specified preference of a prospective consumer is mapped to a vector $1 \times \boldsymbol{K}$ vector X_k. This vector enables the classification of the consumer into one of the $\boldsymbol{C}$ classes using $\boldsymbol{\gamma_c}$ vectors. Once the consumer is classified into a class $\boldsymbol{c}$, the parameters α_c is used along with the X_k vector to calculate the weight all reviews in that class. Based on the weights, the reviews are re-ranked and shown to the consumer in a descending order.

3.3 Online Review Personalization using LCR Model

This section discusses the use of the LCR model developed in Section 3.2 to personalize the ranking of the reviews for consumers at run time. There are two steps in the process that include finding out the class of a consumer based on his/her product feature preferences (see Section 3.3.1) and using the corresponding

regression model of the consumer class for re-ranking the reviews and presenting them to the consumer (see Section 3.3.2).

Figure 11. Example of specified preferences from Amazon.com ORP

3.3.1 Identification of Prospective Consumer's Class based on Specified Preferences

Every ORP provides a way for helping prospective consumers narrow down the products that they are looking to buy. For example, a prospective consumer looking for a cell phone to buy at Amazon.com is presented with the options (see Figure 11) to narrow down the cell phone models that are shown him/her. FDPPR proposes to use such kind of mechanism in ORPs to capture prospective consumer's specified product feature preferences to display personalized review rankings.

It is assumed that a prospective consumer is a utility maximizing consumer. So, he/she has a utility function that helps him/her decide what to buy based on the information he/she collects. It is also assumed in this research that, the utility function comprises of "m" product features; and at any given time, the consumer can show interest in "p" out of the "m" features ($p <= m$).

When a prospective consumer has not specified preference for any features, i.e., $p=0$ information state, he/she is presented with a universal ranking of reviews (i.e. the baseline scenario). However, when a prospective consumer chooses one (i.e., $p= 1$)

product feature preference (e.g., internal memory), reviews are regrouped/selected and re-ordered in a way that is personalized according to his/her specified product feature preference. The range of p for a prospective consumer will be [0, m]. Also, each of these product features usually has multiple levels. For example, internal memory can be 4 GB, 8 GB, 16 GB, 32 GB, and 64 GB. These multiple levels make it very difficult to have different ranked lists of reviews for each possible combination (e.g., in Figure 11 there will be 5 * 6 * 5 * 6) of choices that a prospective consumer can make. So rather than examining all possible combinations, this research creates q lists (i.e. latent classes) based on different combinations of p choices and their levels using latent class regression models. This research uses parsimony-based indices [130] (e.g. Bayesian Information Criterion, or BIC) to identify an optimal number of latent classes. When p=0 (the current state), no "personal choices or preferences" are incorporated to rank product reviews. When p >= 1, this research incorporates prospective consumer product feature preferences and the consumer are shown with only ordered reviews that he/she is interested in.

Once the consumer selects a particular product feature (i.e. p >= 1), the LCR model is used to identify which class he/she belongs to. The identification of class belongingness is important as reviews are ranked based on the regression equation that describes the consumer class of the targeted consumer.

3.3.2 Re-ranking Reviews based on Prospective Consumer's Class

Once the particular *consumer class* of a prospective consumer is determined, the regression model developed for that *consumer class* is used to recalculate the

helpfulness score of all reviews. These new scores are used to re-rank reviews in a decreasing order of relevance (i.e., personal helpfulness scores). As the regression models are already developed in the offline analysis stage, the calculation of personal review helpfulness can be done online. When new reviews are added, the regression models can be updated to keep the LCR model up-to-date.

Chapter 4. A Case Study - Building a Review Ranking Personalizer using FDPPR

One of the critical steps in this research is to understand the products and the associated reviews so that the review ranking personalizer can be built using the product features that influence personalization. Towards that end, a large number of reviews spanning across multiple products were collected for the offline review analysis which in turn helps in the development of online review ranking personalizer. A review scrapper that can collect data from Amazon.com was developed in this research to automatically collect and extract data for product review analysis. The Amazon review collector collected more than 90,000 reviews spanning 99 products. The products were selected from three categories, including electronic, home, and sports goods to study review characteristics.

Data collection, extraction, consolidation and storage framework are described through multiple steps in Section 4.1, which were used for offline review analysis. Section 4.2 describes the development of the online review ranking personalizer based on the curated review data. While describing details of each step, an example case (Amazon product id: B00KHR4ZL6 for Microsoft Surface Pro 3) is provided with real outputs of the analysis in each step.

4.1 Data Collection and Pre-Processing for Offline Review Analysis

For development of the FDPPR framework, a review corpus spanning multiple products was created. In many previous studies on review helpfulness prediction, the

researchers have analyzed product reviews collected from Amazon.com [11, 22, 24, 42, 45, 50, 51, 131]. In this research, reviews from multiple products were also downloaded from Amazon.com using a custom review crawler. Once the product review data were collected from Amazon.com, they were processed to extract product features, stylistics characteristics (i.e. textual and meta-characteristics) and polarity information (i.e. sentiments) from the reviews. The details of these text extraction and processing pipeline are presented next.

4.1.1 Review Corpus Creation

The first step of this research was to collect reviews from an ORP. For the research, multiple product reviews from Amazon.com website were extracted from Amazon.com ORP. The software designed and developed for this task was called Amazon Review Collector (ARC). ARC (see Figure 12) was a web scrapper that was developed to automatically collect all the product reviews for any given product and their associated structural information such as star rating, number of comments etc. from Amazon. ARC was developed using Java and uses open source tools like Apache Tika library[6], JSoup Parser[7], Java XML parser and XPath Query language[8] to collect raw reviews and associated metadata.

Every product that is sold on Amazon.com has a unique product identification number (e.g. B00KHR4ZL6 for Microsoft Surface Pro 3). The Amazon ORP that supports product review writing, filtering, and feedback process is also tied to the

[6] http://tika.apache.org/
[7] http://jsoup.org/
[8] http://docs.oracle.com/javase/7/docs/api/javax/xml/xpath/package-summary.html

unique identification number of the product (e.g. https://www.amazon.com/product-reviews/B00KHR4ZL6/ for the review of the product B00KHR4ZL6). A list of Amazon product identification numbers was manually collected for this research by visiting different product categories on Amazon.com (i.e. https://www.amazon.com/gp/site-directory/). The collected Amazon product identification numbers were stored in a MySQL[9] database server. ARC was configured to read the Amazon product identification number from the database and go online to collect the details of the products and download all the associated reviews from Amazon.com. ARC was designed to crawl multiple pages automatically and get all the reviews for a product by automatically creating the paginated links. A front-end user interface was created to monitor and analyze the review data collected from Amazon.com.

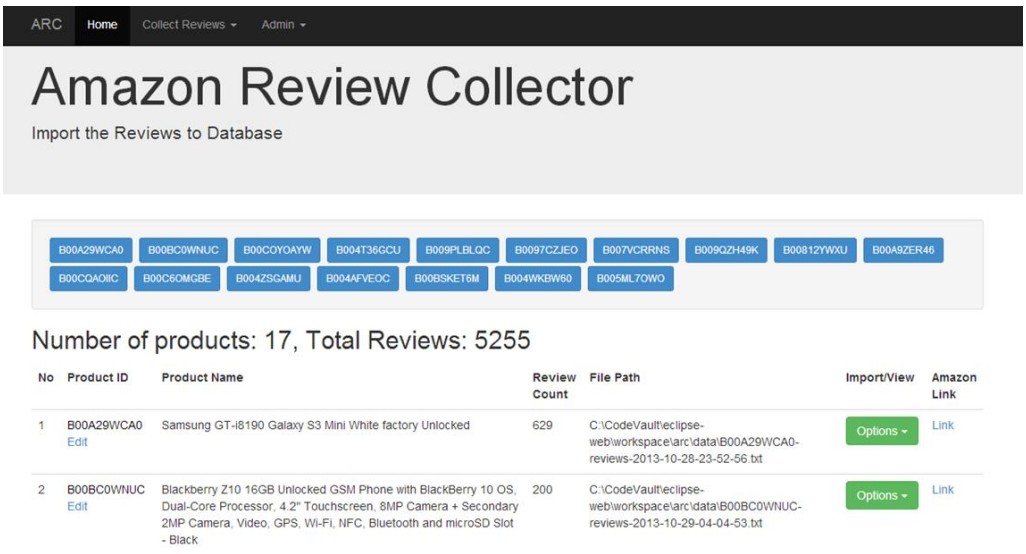


Figure 12. ARC to crawl and extract review data from Amazon

[9] https://www.mysql.com/products/community/

Figure 13 shows the ARC review extraction pipeline. The program takes two parameters as inputs, namely the unique amazon product identification number and the number of review pages for the product. Based on the amazon product identification number, the product description page was automatically crawled. After this, ARC crawled each review page to extract review information and store it locally.

Table 4. Steps for crawling and collecting reviews from Amazon.com

Step	Details	Example
1	Get the amazon product identification number	B00KHR4ZL6
2	Get the total number of pages of reviews present to support paginated review collection	46
3	Get details about the product by crawling the product page	https://www.amazon.com/ dp/B00KHR4ZL6/
4	Collect the reviews for the product by crawling all the review pages	https://www.amazon.com/ product-reviews /B00KHR4ZL6/ ?pageNumber=1
5	Store the collected review data on disk	

Table 4 shows the steps of the review collection process for a single product (e.g. B00KHR4ZL6 for Microsoft Surface Pro 3). The product page at Amazon.com is always present at https://www.amazon.com/ dp/<product identification number>/ (i.e. https://www.amazon.com/ dp/B00KHR4ZL6/) and the reviews are present at https://www.amazon.com/ product-reviews /<product identification number/ ?pageNumber=<page number> (i.e. https://www.amazon.com/ product-reviews /B00KHR4ZL6/ ?pageNumber=1). In this case, there were 46 pages of review present. The ARC visited every page from https://www.amazon.com/ product-reviews /B00KHR4ZL6/ ?pageNumber=1 to https://www.amazon.com/ product-reviews /B00KHR4ZL6/ ?pageNumber=46 to download the reviews and stored them

locally in one single text file. This text file also stored meta-information of reviews such as the star rating, number of likes, and number of comments, etc.

Figure 13. ARC used for extracting reviews

The Apache Tika library was used to extract a web page from the Amazon.com website. Once a webpage was downloaded, the HTML content in that web page was parsed using the JSoup library. JSoup library is a HTML parser that provides helpful methods to standardize the content of the HTML file so that it can be queried as a XML document.

Product ID	B00KHR4ZL6
Review Title	Surface Pro 3 is the Future
Review Text	I currently own Surface 2, MacBook PRO 15 inch, and an iPhone 5 so I am not bias to either side (replacing the MacBook Pro/S2 with the SP3). Choose the product that is best featured, not best brand - this drives competition and consumer quality. I am goi...
Reviewer Name	C. vu
Comments	23
Date of Review	6/20/2014 0:00
People Found Useful	672
People Total	721
Star Rating	5

Figure 14. Amazon Review Crawler: output from review extraction

Once the HTML file was standardized as XML, XPath queries were executed against the review content to extract individual reviews from the review page. XPath is a query language that is used to navigate through elements and attributes in an XML

document. Each review page from Amazon contains up to 10 individual reviews as well as their meta information. By querying the data using XPath each individual review along with their meta characteristics were extracted and stored locally for later processing. Figure 14 shows one automatically extracted review for the Microsoft Surface Pro 3 laptop in a text file.

Table 5. Sample list of product reviews collected from Amazon

No	Amazon ID	Product Description	Reviews
1	B00BGO0Q9O	Fitbit Flex Wireless Activity + Sleep Wristband	12440
2	B007JR532M	SanDisk Cruzer CZ36 32GB USB 2.0 Flash Drive	5700
3	B00DHJ8QLQ	Brother MFC-J870DW Wireless Color Inkjet Printer with Scanner	3220
4	B00008Y0VN	Celestron SkyMaster Giant 15x70 Binoculars with Tripod Adapter	3180
5	B005GK3IVW	iRobot Roomba 770 Robotic Vacuum Cleaner	2520
6	B001EJMS6K	Iron Gym Total Upper Body Workout Bar	2490
7	B00KHR4ZL6	Microsoft Surface Pro 3 PS2-00001 12-Inch Pro 3	1041
8	B006ZP8UOW	Foscam FI8910W Pan & Tilt IP/Network Camera with Two-Way Audio and Night Vision	2135
9	B001ARYU58	Bowflex SelectTech 552 Adjustable Dumbbells (Pair)	2000
10	B002DW92IE	Monster iCarPlay Cassette Adapter 800 for iPod and iPhone -3 feet	1965

Using ARC more than 90,000 reviews spanning across 99 products from Amazon reviews were collected. The reviews were posted between November 2000 and August 2015. A sample list of products and their associated number of reviews is presented in Table 5. In the early stages of this research, all these reviews provided deep insights into the characteristics of the reviews as well as the parameters that could be used for the development of the review ranking personalizer as part of FDPPR framework.

Once the reviews of a product were downloaded and parsed, they were stored in the MySQL database which served as the baseline review corpus for the research.

4.1.2 Review Meta Characteristics Extraction

Each element on the product page and review page on Amazon ORP could be extracted using XPath expressions.

Table 6. One Sample Product Review Meta Characteristics Extraction

	XPath Expression	Extracted Feature
List of Review ids	.//*[@id='cm_cr-review_list']	R1GPD028CILB07, R2OALNG029OIYL, etc.
Customer rating for a review	.//*[@id='customer_review-R1GPD028CILB07']/div[1]/a[1]/i/span	5.0 out of 5 stars
Number of people found useful	.//*[@id='customer_review-R1GPD028CILB07']/div[5]/div/span[3]/span/span[1]/span	25
Date of review	.//*[@id='customer_review-R1GPD028CILB07']/div[2]/span[4]	On June 5, 2014
Number of comments	.//*[@id='customer_review-R1GPD028CILB07']/div[5]/div/a/span/span[2]	1 comment

XPath expressions can query any node in a XML document to get the value of that node element and attribute as needed. Table 6 shows a sample of XPath queries that

were executed to get several meta characteristics from a single review. The results of the queries were stored in the database for further analysis.

The same technique was used to extract the text characteristics of a single review as explained in the next section.

4.1.3 Review Textual Data Extraction

Similar to extracting review meta-characteristics (see Section 4.1.2), the textual characteristics of reviews were collected using XPath queries. The result of extracting a single review data is presented in Table 7.

Table 7. One Sample Product Review Collected from Amazon.com

Product id	B00KHR4ZL6
Review title	Surface Pro 3 is the Future
Review text (truncated)	I currently own Surface 2, MacBook PRO 15 inch, and an iPhone 5 so I am not bias to either side (replacing the MacBook Pro/S2 with the SP3). Choose the product that is best featured, not best brand - this drives competition and consumer quality. I am going to list PROs and CONs organized into significant and common. Significant PROs/CONs is a term for uniqueness or high-end / setback or low-end. Common is a term for similar features or specifications amongst its peers. Most comparisons are between the SP3 12" to MacBook Air 11.6" (2013) model (due to its public familiarity, size, and prestige). The MacBook Air 11.6" (2014) model improves with a slight faster CPU clock (0.1 GHz) and PCIe SSD almost making it the same MacBook Air - refer to TechnoBuffalo (2014) for technical specifications of the MacBook Air 11.6" 2014 model, thus it is safe to compare SP3 with the MacBook Air 2013 model. Everything will be cited (loosely APA format) for viewers to check credibility.
Reviewer name	C. Vu

The review text data was collected from all the reviews of a product. Once a review text data was collected, the characteristics of the text were extracted using natural language processing. For this task, the Stanford NLP software tool [132] was used. The extracted details from a review text are presented in Table 8.

Table 8. Product Review Textual Characteristics	
Product id	B00KHR4ZL6
Review title	Surface Pro 3 is the Future
Review text with no stop words (truncated)	I currently Surface 2, MacBook PRO 15 inch, iPhone 5 I bias (replacing MacBook Pro/S2 SP3). Choose product best featured, best brand - drives competition consumer quality. I going list PROs CONs organized significant common. Significant PROs/CONs term uniqueness high-end / setback low-end. Common term similar features specifications peers. Most comparisons SP3 12" MacBook Air 11.6" (2013) model (due public familiarity, size, prestige). The MacBook Air 11.6" (2014) model improves slight faster CPU clock (0.1 GHz) PCIe SSD making MacBook Air - refer TechnoBuffalo (2014) technical specifications MacBook Air 11.6" 2014 model, safe compare SP3 MacBook Air 2013 model. Everything cited (loosely APA format) viewers check credibility.
Review raw word count	2111
Review word count after removing stop words	1383

4.1.4 Product Feature Extraction

One major step of the research is to identify product features that are discussed in the reviews of that product. In this research, topic modeling using LDA was used to identify the product features that are discussed the most by the reviewers. A basic LDA analysis using the Mallet library[10] was performed for product feature extraction. The code took individual product reviews as input and extracted pre-defined number

[10] http://mallet.cs.umass.edu/

of topics from reviews (see Figure 15) along with the top keywords appeared under each of the topics. Once the topics and associated top keywords were extracted from the reviews, the product features appeared most in these topics were used for the development of the review ranking personalizer.

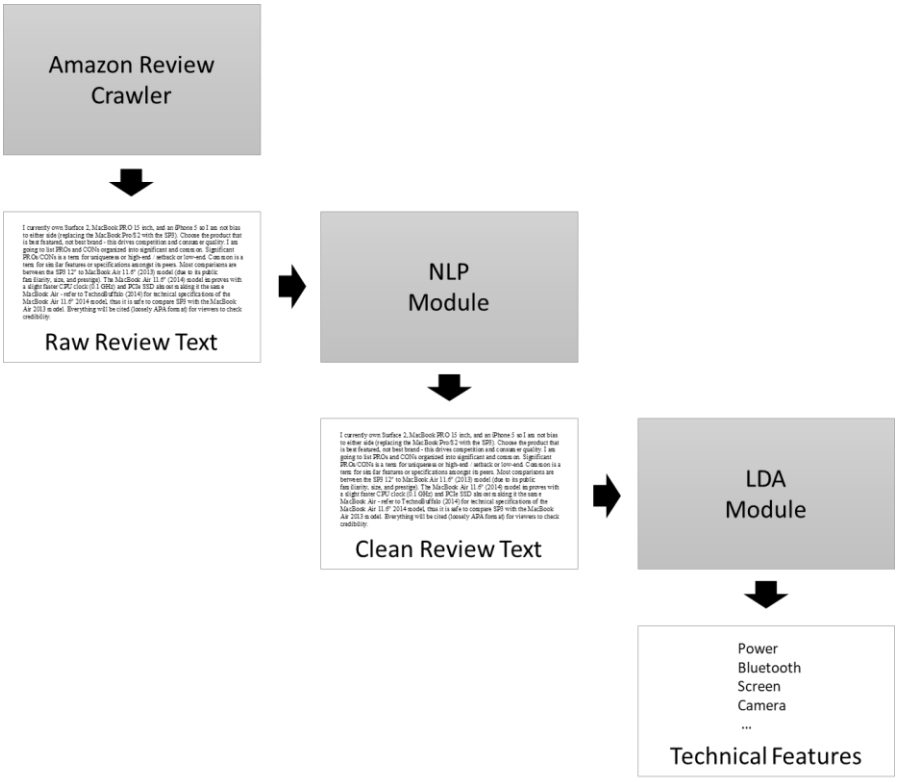


Figure 15. Product feature identification using LDA from Amazon reviews

In particular, the following steps were involved in the LDA process on the product reviews.

Step1: Extraction of Review Text into Separate Text Document

As discussed in Section 4.1.3, the review text data were first collected from Amazon ORP and saved in a MySQL database after removing the stop words from the collected data. For analysis using Mallet library, the cleaned review text data (i.e.

reviews with no stop words) were placed in separate text files. For the Microsoft Surface 3 review analysis, 1041 separate text files were created where each text file represented one review. Next, these text files were imported into mallet's internal format.

Step2: Importing Reviews into Mallet's Internal Format

Using mallet for LDA analysis requires that the data should be in the internal mallet format. It represents data as lists of "instances". All mallet instances include a data object. An instance can also include a name and (in classification contexts) a label. Each review text was considered as a single instance for LDA analysis using Mallet.

Table 9. Importing individual reviews into Mallet internal format

```
C:\> mallet import-dir --input "C:\>Data\B00KHR4ZL6" --
output B00KHR4ZL6.mallet
```

Table 9 shows the command that was executed to import all the exported reviews into mallet format for Amazon product id B00KHR4ZL6 (i.e. Microsoft Surface Pro 3). The LDA model was built by using the imported mallet file.

Step3: Setting Parameters and Building LDA Model

The next step was to set various parameters before building the LDA model. The parameters required are described below. The parameters that were set before doing the LDA analysis are given below. The default values (most common values used in many previous research) were given as the parameters.

- --*input [FILE]*: specifies the MALLET collection file created in the previous step.

- --*num-topics*: Number of topics (20), given by researcher
- *num-iterations*: The number of sampling iterations is a tradeoff between the time taken to complete sampling and the quality of the topic model.
- *optimize-interval*: This option turns on hyper parameter optimization, which allows the model to better fit the data by allowing some topics to be more prominent than others. Optimization is performed every 10 iterations.
- *optimize-burn-in*: The number of iterations before hyper parameter optimization begins. Default is twice the optimize interval.
- *alpha_sum*: This is the magnitude of the Dirichlet prior over the topic distribution of a document. The default value is 5.0. With 10 topics, this setting leads to a Dirichlet with parameter $\alpha_k = 0.5$. It is the number of "pseudo-words", divided evenly between all topics that are present in every document no matter how the other words are allocated to topics.
- *beta*: This is the per-word weight of the Dirichlet prior over topic-word distributions. The magnitude of the distribution (the sum over all words of this parameter) is determined by the number of words in the vocabulary. Again, this value may change due to hyper parameter optimization.

Table 10. Building LDA model using Mallet internal format

```
C:\> mallet train-topics --input B00KHR4ZL6.mallet --
num-topics 20 --optimize-interval 20 --output-topic-
keys B00KHR4ZL6_top_keywords.txt --output-doc-topics
B00KHR4ZL6_topic_percentage.txt
```

Table 10 shows a sample command to build the LDA model as well as the output of the LDA analysis in a text file (e.g. B00KHR4ZL6_top_keywords.txt) for interpretation.

Step4: Product Feature Extraction by Interpreting the Results of the Analysis

LDA helps with identifying the important product features that are described in the product reviews. Table 11 shows a sample output from the LDA model built from the Microsoft Surface Pro 3 reviews. The table shows many product features that contributed towards the identified topics. These product features were incorporated to build the review ranking personalizer. A manual supervision was performed to select the top terms appearing under each topic to be considered as potential product features. For example, it can be seen that from Table 11, Topic 0 is mostly about display resolution and screen factors of the product. These product features were used as the covariates in the LCR analysis as described in section 4.2.

Table 11. Product feature extraction from top terms of LDA topics

Topic		Top Terms in a Topic	Selected Terms
0	0.00796	**display** works **touch** **resolution** games computing time	display, resolution, screen, touch
1	0.05398	**windows apps** android system download **app** installed issues software google	app, apps
2	0.03655	**power** onenote hrs day kindle life computer day **battery**	battery life, power
3	0.25923	**keyboard** click device **mouse** allows laptop issues **button** wireless	keyboard, mouse, button
4	0.01096	**stylus** perfect **write** smooth **pen** users android cloud	pen, stylus
5	0.03975	**port** power surface **usb** firmware cover time **wireless** cloud	port, usb

4.1.5 Review Polarity Extraction using Sentiment Analysis

As discussed earlier in section 3.2.4, for sentiment analysis, an opinion lexicon list developed by Finn Årup Nielsen [133] was used to assign a positive score and a negative score to every review. As a part of this research, a custom java code (i.e. sentiment analysis module) was developed to tokenize the words in the cleaned review text.

Once the review was tokenized, the words were compared against the opinion lexicon (see Figure 16). When the words matched the opinion lexicon, the score of the word was added to the positive and negative score of the review.

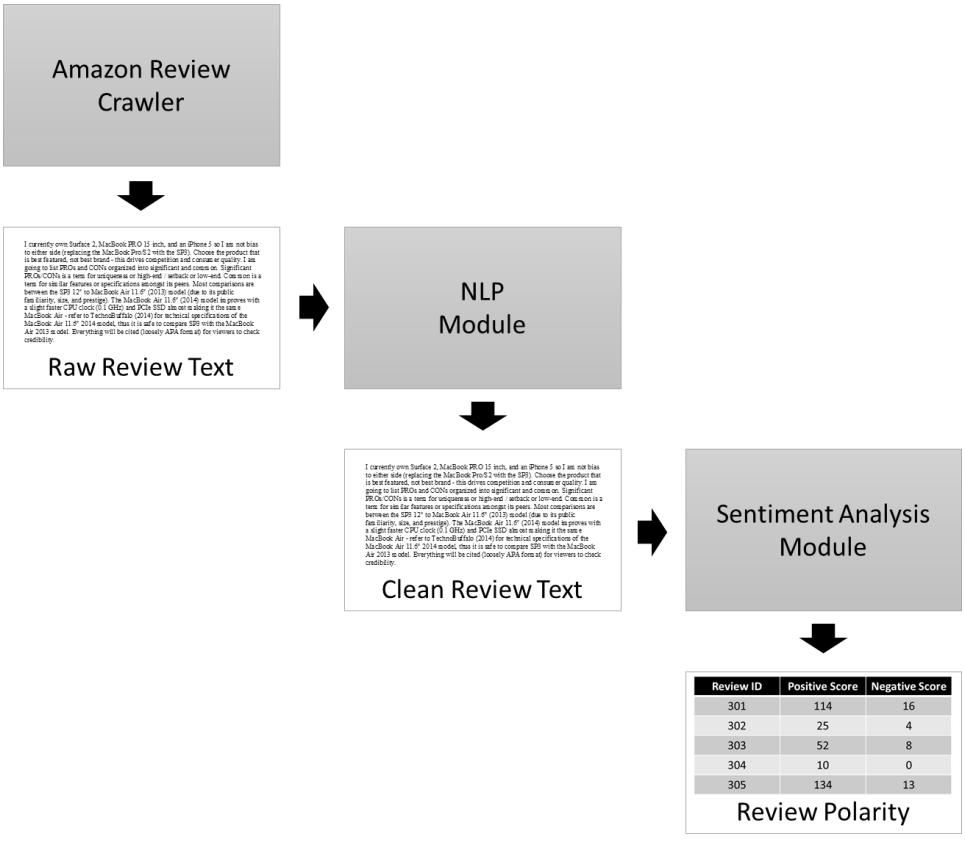


Figure 16. Sentiment analysis on Amazon reviews

This process was continued until there were no more words left in the review. At the end of the process, a cumulative positive and negative score of the review was saved in the database and was used towards the development of the review ranking personalizer.

Table 12. Review sentiment score calculation		
Review id	Positive Score	Negative Score
1026	114	16
1027	25	4
1028	52	8
1029	10	0
1030	134	13
1031	18	3
1032	6	4

Table 12 shows the calculated positive and negative scores for the corresponding reviews.

Once the meta characteristics, textual characteristics, product features, and sentiment scores were extracted from the product reviews, the LCR model was built to develop the online review ranking personalizer.

4.2 Development of an Online Review Ranking Personalizer

Development of the proposed online review ranking personalizer mainly involved the analysis of product reviews with the LCR model to find out distinct consumer classes and generate a separate review ranking for each consumer class. After applying data pre-processing steps as described in section 4.1 on the reviews of the product (Microsoft Surface Pro 3), the LCR model was applied on the preprocessed review data to develop a review ranking personalizer system for that product. As described in

section 4.1, reviews of the product were extracted from amazon.com and pre-processed to extract various review characteristics (review stylistic characteristics and sentiments) and product features discussed in the reviews. There were a total of 1041 unique rows in the review dataset where each row represented an individual review.

Table 13. Basic Descriptive Statistics

	Variables	Description	Mean	Stdv
Dependent variable (DV)	People found useful	Indicates the helpfulness of the review	19.42	101.60
Independent Variables (IV)	Number of days	It specifies the how long has been the review posted since the first review.	187.23	93.28
	Word count	Review text with no stop words and special characters (i.e. length of the review)	638.58	327.70
	Total Number of people	Specifies the total number of people that have read the review	22.65	106.48
	Comments	Specifies the number of comments the review has got.	186.03	232.83
	Rating	The number of stars the review has got	4.05	1.44
	Positive score	Positive sentiments of review	18.62	21.81
	Negative score	Negative sentiments of review	5.62	7.85
Covariates (CV)	Pen, stylus	Product features mostly discussed in the reviews	NA	NA
	App, apps		NA	NA
	Battery life, power		NA	NA
	Display, resolution, screen, touch		NA	NA
	Keyboard, mouse, button		NA	NA
	Port, usb		NA	NA

And each review was associated with different review characteristics and product features. Basic descriptive statistics were obtained as shown in Table 13. The table also describes each variable used in the LCR analysis.

4.2.1 Building and selecting the LCR analysis model

In this research, as described in section 4.1, Latent GOLD[11] software was used to perform LCR analysis of the review dataset. LCR analysis was performed on the review data by taking people found useful as the dependent variable (DV) in the analysis. The predictor variables or the independent variables (IV) were used to predict the helpfulness of the reviews while the identified product features were used as covariates to predict the class membership of the reviews as shown in Table 13.

Table 14. LCR Model Selection

		LL	**BIC(LL)**	p-value	**R^2**
Model 1	1-Class Regression	-2276.99	4611.436	3.0e-605	0.8924
Model 2	2-Class Regression	-1394.48	2948.544	2.70E-289	0.9974
Model 3	**3-Class Regression**	**-1265.55**	**2792.815**	**1.90E-250**	**0.9994**
Model 4	4-Class Regression	-1232.51	2828.889	1.60E-244	0.9994
Model 5	5-Class Regression	-1226.52	2919.035	7.80E-248	0.9995

In LCR analysis, various regression models were developed for a different number of classes (e.g. 1 class, 2 class, 3 class and so on) and the optimum model among them was chosen by looking at various model statistics such as Akaike information criterion (AICs), Bayesian information criterion (BICs) and R^2 values. In particular, a latent class model with a lower BIC and a higher R^2 (overall) value would be

[11] http://www.statisticalinnovations.com/latent-gold-5-1/

selected. In LCR analysis, each model was also associated with two underlying output models, namely, (i) *model for dependent* (i.e. regression analysis) and (ii) *model for classes* (i.e. covariates statistics). After choosing the optimal model, the coefficients of the underlying associated model were used to determine the class membership of the covariates and estimate a regression equation for each of these classes. In this case, after running the LCR models for different classes, the best model was model 3 with a lower BIC (2792.8154) and higher R^2value (0.9994) for this review dataset (see Table 14).

4.2.2 Determining class belonging of each co-variate from the output of LCR model (*model for classes*)

The first output of the LCR model was the *model for classes*. As shown in Table 15, every covariate had a certain degree of class membership in each of the three classes. The degree of class membership is shown by the coefficient values as shown in the table.

Table 15. Model for classes

	Class1	Class2	Class3	Wald	p-value
Intercept	2.0912	-0.6013	-1.4899	174.3225	1.40E-38
Covariates	Class1	Class2	Class3	Wald	p-value
app, apps	0.0347	-0.1997	0.1649	2.2806	0.32
battery life, power	-0.1856	0.2035	-0.0179	10.5699	0.0051
display, resolution, screen, touch	-0.0257	-0.0601	0.0858	2.1011	0.05
keyboard, mouse, button	0.0096	-0.0614	0.0518	0.675	0.02
pen, stylus	-0.2428	**0.4232**	-0.1804	11.2993	0.0035
port, usb	0.2399	-0.3764	0.1365	6.6859	0.035

Those coefficients were used to calculate each covariate's class membership. For example, pen~stylus had more contribution towards class2 (beta ~ 0.4232) than class1

(beta~ -0.2428) and class3 (beta~ -0.1804). Also, the p value associated with each covariate showed that some covariates were highly significant (i.e. $p < 0.05$) in predicting the class membership such as pen/stylus, port/usb, and battery/power. Those highly significant covariates were then considered as important product features of the product on which consumers have mostly discussed in the reviews.

The *model for classes* output was used to find the consumer class of a consumer based on his/her selection of a covariates (i.e. preferred product features). For example, if a consumer selected a product feature, such as "pen, stylus", then he/she would primarily belong to class2 as the pen/stylus had mostly contributed to class2. He/she would also have a membership with class 1 and class 2, however with a lower degree. In this research, a participant was assigned to one particular class only based on his/her selection of preferred product feature(s). However, LCR analysis allows probabilistic assignment of the same individual to multiple classes and uses a weighted average technique to recalculate the scores of reviews across multiple classes and re-rank reviews per each class. That will be explored in future research given the scope of this study.

4.2.3 Estimate the regression models for each of the classes and re-calculate and re-rank the helpfulness of all the reviews.

The second output of the LCR model, model for dependent is shown in Table 16. The overall R^2 for the classes was 0.9994 indicating the result was statistically significant. The result shows that, there was significant differences among the predictors' contribution toward the helpfulness of the reviews on all three classes ($p < 0.05$), except positive score ($p > 0.05$) which had more or less same significance on all three

classes. While the positive coefficients were positively contributing toward the helpfulness of the reviews and negative coefficients were negatively influencing the helpfulness of the reviews. However, the amount of contribution of each predictor (i.e. coefficients of each predictor) was different in each class. For example, it can be interpreted that, class 1 is mostly about word count and positive rating and class1 is positively influenced by the non-stopword word count but negatively influenced by product rating. It indicates that consumers belonging to a particular class may emphasize different characteristics of reviews.

Model for dependent is used to calculate the helpfulness of the reviews for each class. More specifically, the regression coefficients of the predictor variables in each class are used to calculate the helpfulness of the all reviews for that class. For each class, the steps for calculating the helpfulness of the reviews are given below.

1. Multiplying the coefficients of each of the predictors into the corresponding value in the dataset to estimate the regression equation for each of the classes which can be represented as the following regression equation:

   $Helpfulness_c = Constant_c\ (intercept) + \beta 1_c\ word\_count + \beta 2_c\ comments + \beta 3_c\ num\_days + \beta 4_c\ negative\ score + \beta 5_c\ total\_num\_people + \beta 6_c\ positive\_score + \beta 7_c\ rating + \varepsilon$

2. Calculating helpfulness of all the reviews for each class using the corresponding regression equation.

In this case study, for each class, a separate regression equation was developed using the corresponding regression coefficients. Then helpfulness scores of all the reviews

were calculated for each class using the corresponding regression equation and reviews were ranked based on the calculated scores. In other words, each class had a separate review ranking that showed reviews to the consumers in a different order than other classes.

Table 16. Model for dependent				
	Class1	**Class2**	**Class3**	
R^2	0.7509	0.9985	0.9998	
people found useful	Class1	Class2	Class3	p-value
Intercept	0.6795	1.3828	1.8205	5.70E-60
Predictors	Class1	Class2	Class3	p-value
Word count	0.0076	0.0042	0.0007	3.70E-05
Comments	0.0643	0.0587	-0.0699	1.30E-12
Number of days	0.0043	0.0082	0.0096	4.60E-103
Negative score	-0.0141	-0.0411	0.0244	7.50E-10
Total Number of people	0.0037	0.0023	0.0046	1.20E-27
Positive score	-0.0038	0.0041	-0.0024	0.46
Rating	-0.1772	-0.2296	-0.1619	1.30E-22

Table 17 shows an example of LCR based helpfulness scores calculated for all 3 classes, for 5 random reviews of the product used in this case study. It can be seen that, LCR based helpfulness scores of reviews (i.e. Class 1 score, Class 2 score and Class 3 score) were different in each class and upon ranked by the scores of each class, it gave three separate rankings of the reviews, one for each class. Also, the given example showed that the helpfulness scores of the reviews were zero in the original helpfulness ranking. However, after applying LCR analysis on the reviews, the helpfulness scores were changed to non- zero scores. This indicates that LCR analysis calculates helpfulness scores of all the reviews of a product including the reviews that have not got any helpfulness score in the original helpfulness ranking. For example, Figure 17 shows an excerpt from a review that was ranked high by LCR

ranking for consumer class 2 of product2 (i.e. Microsoft Surface pro tablet). Consumers that belong to class 2 preferred “pen/stylus” as the product feature.

> This comes with only the Surface Pro 3 with the pen. ... Overall it's nice tablet. I'd buy a cheap glove or glove liner for using the pen. Because it's a tablet, it will detect the side of your hand on the screen while you're writing. This causes extra pen marks from the side of your hand. ...

Figure 17. A Review discussing about the pen feature of tablet

The original people found useful score (i.e. the helpfulness score) for the review shown in is zero in the helpfulness ranking, but the content of the review suggests that the review will be useful for a consumer trying to know more about the “pen/stylus” feature of the tablet. This example shows that LCR ranking can help bring a relevant review to the notice of prospective consumers by weighing them higher in the ranking system.

Table 18 shows the reviews ranked in three different order (LCR based) including the original helpfulness based order. When a consumer selected a product feature, such as pen/stylus, he/she was first classified to class2 (as described in section 4.2.3) and then class 2 review ranking was shown to him/her where reviews were ordered in differently than reviews ordered in class1 and class3.

Table 17. Example LCR based helpfulness of the reviews											
id	Clean word count	Comments	Number of days	Negative score	People found useful	People total	Positive score	Rating	Class 1 score	Class 2 score	Class 3 score
639	140	0	29	6	0	0	15	5	0.171	0.4849	1.3862
640	125	0	4	4	0	0	18	4	0.2083	0.5788	1.2524
641	133	0	0	0	0	0	17	5	0.1709	0.5313	0.9685
648	103	0	12	3	0	0	11	5	0.1082	0.4335	1.1725
650	106	0	210	2	0	0	19	5	0.9229	2.1184	3.0246

Table 18. Reviews ranked by different order				
Reviews	Helpfulness rank	LCR Class1 rank	LCR Class2 rank	LCR Class3 rank
639	612	612	615	615
640	613	613	627	627
641	614	615	614	614
648	615	627	613	613
650	627	614	612	612

Chapter 5. Performance Evaluation of FDPPR

This chapter presents an evaluation of the proposed FDPPR framework for providing personalized review ranking to the consumers. Section 5.1 presents hypotheses development that are aimed at comparing the existing review ranking with the review ranking developed by FDPPR framework. Section 5.2 presents research methodology of the online user study in detail. Section 5.4 discusses the experimental procedure followed by Section 5.3 that presents details about the participants of the study.

5.1 Hypotheses Development

In this research, an online user study was conducted to evaluate the effectiveness of the proposed FDPPR framework by comparing the performance and user perceptions of the proposed product feature oriented, personalized review ranking with the most widely used helpfulness ranking. In particular, it was tested whether LCR ranking that takes consumer's personal product feature preferences into consideration, provides more relevant and helpful reviews than traditional helpfulness ranking. The performance and user perceptions are captured by using 3 widely used measures in the IS research for system evaluation, namely, (1) relevance, (2) knowledge and (3) satisfaction. A website (i.e. www.curetext.com) was designed to perform this online experiment.

5.1.1 Personalized Review Ranking vs. Helpfulness Review Ranking with respect to Relevance

According to various models of information systems (IS) success, information *relevance* has been an important dimension of information quality [104]. A study of digital libraries found that relevance of the information retrieved from the information retrieval (IR) systems had a significant effect on perceived helpfulness of IR systems [134]. Helpfulness ranking provides a lot of information about a product irrespective of any specific feature(s) of the product. But most of the time consumers have varying interests on different features of a product while evaluating the quality of the product. Thus, providing information about specific features of a product that are of interest to consumers seems more helpful to them. Extant research has shown that providing relevant information to the users based on their preferences (i.e. personalized approach) helps them quickly assess the quality of the items and subsequent decision making [135, 136]. As the amount of reviews abundantly increase with time, it is not only important to find relevant reviews but also presenting them to the consumers in an effective order (i.e. rank) so that users can quickly evaluate the product based on their personal preferences. There have been many studies that have substantiated the need of ordering of information while accessing information because it helps faster judgement and decision making by human beings [137, 138]. Based on the discussed theoretical ground, it is predicted that, the personalized review ranking proposed in this research will provide more relevant information to the consumers than the currently available helpfulness ranking. The first hypothesis is proposed as follows:

Hypothesis 1(H1)*: Consumers will be able to access* ***more relevant reviews*** *from the top reviews ranked by the personalized review ranking (i.e., LCR ranking) as compared to the top reviews based on helpfulness ranking*

5.1.2 Personalized Review Ranking vs. Helpfulness Review Ranking with respect to Knowledge

As a part of the review data analysis performed during this research, it is observed that the top reviews (i.e. reviews ranked high) in the traditional helpfulness ranking system usually contain a lot of discussion about the product features of the product. Those reviews contain in-depth analysis of some product features and their quality. For example, many reviews compare the product features with other competitive products' product features and provide additional details about the product features that increase the knowledge of consumers on the product features. Before making a purchase decision, usually consumers evaluate the quality of a product based on their preferences for certain features of the product. They prefer to read reviews where they gain more information about their preferred product features. In the helpfulness review ranking, consumer's product feature preferences are not taken into consideration. So, it can be assumed that top reviews in such ranking system provide detailed information of some of the product features on which the consumer might not be interested at all. On the other hand, in the personalized review ranking, the consumer is first requested to provide his/her choice for a particular product feature of the product on which he/she want to see some reviews. Based on that feedback, the reviews are ranked and presented to the consumers where the top reviews in the raking provide more relevant discussion and information about the consumer desired product features. The assumption is that the consumer will prefer the reviews that

provide more knowledge about the product features that is of interest to a perspective consumer. Thus, the second hypothesis is proposed as follows:

Hypothesis 2 (H2)*: The top reviews ranked by personalized review ranking (i.e. LCR ranking) will provide a* ***higher level of knowledge*** *about interested features of a product to consumers than those ranked by helpfulness ranking*

5.1.3 Personalized Review Ranking vs. Helpfulness Review Ranking with respect to Satisfaction

Consumer satisfaction has been considered as a perceptual or subjective measure of system success and one of the important determinants of information system effectiveness [104]. If an ORP makes it very difficult and time consuming for consumers to find reviews that meet their needs, consumers will become dissatisfied and look elsewhere [139, 140] . Traditional helpfulness ranking does not consider consumers' specific needs (i.e., product feature preferences) while assessing and ranking reviews, which may end up ranking those reviews that receive a large number of helpful votes but do not necessarily comment on the specific product features important to a consumer, leading to high consumer dissatisfaction. In contrast, the personalized review ranking better caters a consumer's need by taking product features commented in reviews and consumer preferences into consideration, thus can make the search for relevant reviews much easier for consumers and better satisfy them than the helpfulness ranking. Based on this assumption, the third hypothesis is proposed as follows:

Hypothesis 3 (H3)*: Personalized review ranking (i.e. LCR ranking) will lead to* ***higher consumer satisfaction*** *than helpfulness ranking*

5.2 Research Methodology

To test the hypotheses discussed in Section 5.1, a user study was designed that captured ranking of a set of reviews by the participants based on their product feature preferences. This user study aimed to derive a ranking of 6 reviews by comparing each review to another review (i.e. pairwise comparison of reviews). A pairwise comparison of 6 reviews (i.e., n = 6) provides a set of 15 (i.e. $\frac{n(n-1)}{2}$) review pairs. Thus, 15 review pairs were shown to the participants sequentially in a random order. For each pair, the participants were asked to compare them and select the review that better matches their interest and provide the relative importance of one review to another by answering a set of questions (i.e. providing scores in a Likert scale). Using these scores, a user ranking was derived for each participant (i.e. each participant's own ranking of the reviews) and was used as gold standard benchmark for comparing the performance between two system generated rankings namely (see Figure 18), (i) people found useful ranking (i.e. helpfulness ranking) and (ii) FDPPR ranking based on the LCR analysis (i.e. LCR ranking).

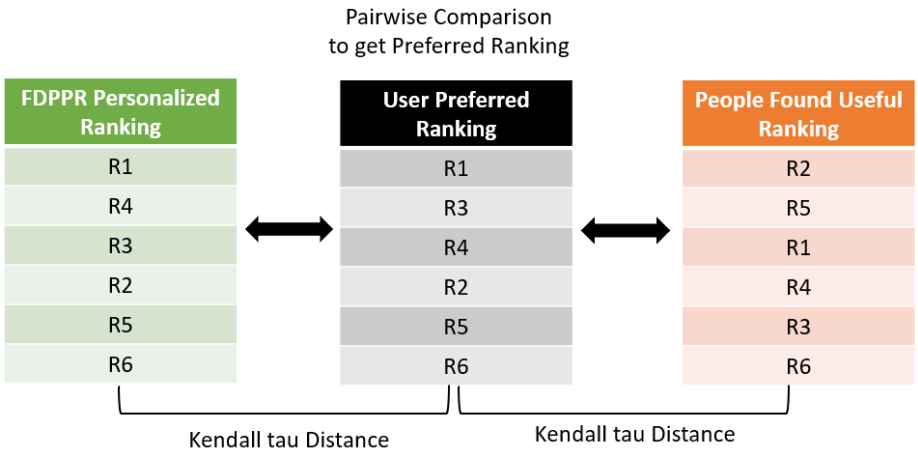


Figure 18. Research methodology for the evaluation of FDPPR using user study

A system generated ranking would be considered better and more desirable if it ranks reviews in a more similar way to a participant's self-ranking.

The design of the user study is divided into five separate steps, namely

(i) Review Selection
(ii) User Class Determination
(iii) User Review Preference Capture using Pairwise Comparison
(iv) Determine the Preferred Ranking of the User
(v) Grade the Review Ranking System

5.2.1 Product and Review Selection

Two products were chosen for the evaluation study that belonged to two different electronic product categories. The first product chosen was Celestron SkyMaster Binoculars with Tripod Adapter (amazon id: B00008Y0VN), referred to as product 1 in the experiment to avoid product bias of the participants, and the second product chosen was Microsoft Surface Pro 3 laptop (amazon id: B00KHR4ZL6, a niche electronic product category), referred to as product 2 (a common electronic product category). In the experiment, along with the 6 eligibility/demographic questions (see Appendix B) and 2 demo video questions (see Appendix B), each participant had to answer 5 questions during each pairwise comparison of reviews. This process resulted into a total of 150 questions (i.e. 15 pairwise comparisons * 2 products * 5 questions) per participant. Thus, to avoid cognitive overload and to limit the number of questions asked to a participant within a reasonable time frame, 6 reviews per product were selected to be ranked by each participant.

Table 19. Review selection process based on two different types of ranking		
Review ID	Reviews by helpfulness ranking	Reviews by LCR ranking per class
1	High	Low
2	High	Low
3	High	Low
4	Low	High
5	Low	High
6	Low	High

Also, among those 6 reviews, to avoid any kind of bias towards a particular type of review, 3 reviews were chosen with high helpfulness scores and very low LCR based scores while the other 3 reviews were chosen to have a high LCR based scores with low helpfulness scores. It is to be noted that while choosing reviews having high LCR scores and low helpfulness score, 3 separate reviews were selected for each consumer class that was identified during LCR model building process. Based on the calculated class of the participant, the corresponding set of reviews were selected for that participant in the user study. Also, to eliminate extremely long or short reviews, reviews of similar length were selected in this user study. More specifically, the number of words in each review would fall in a range of 250 to 350 words.

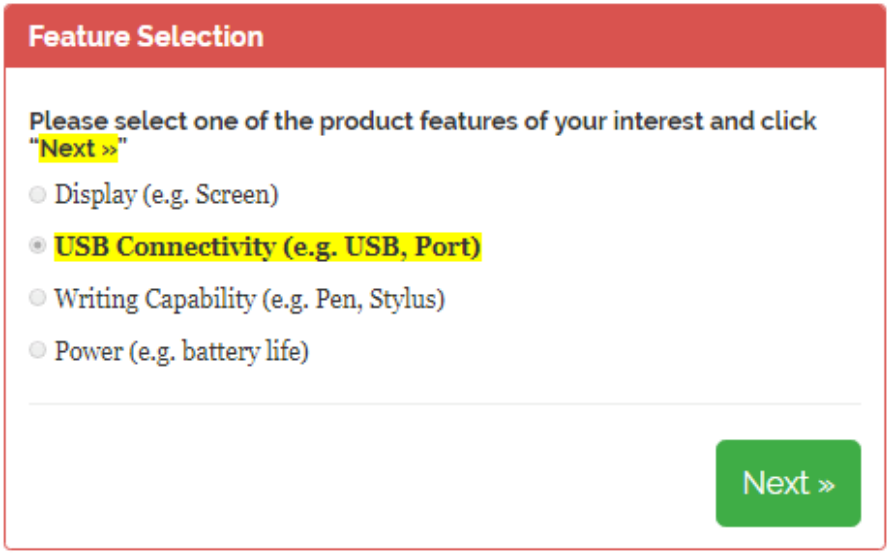


Figure 19. Product feature selection by participant

5.2.2 User Class Determination

In this step, participants were first shown a detail description of a product and were instructed to select a product feature of their choice from a list of available product features as shown in Figure 19. Based on their selected product feature, participants were classified to a particular consumer class pre-determined by the LCR analysis.

5.2.3 User Review Preference Capture using Pairwise Comparison

After a participant selected a particular product feature, he/she was presented with pairwise comparison of reviews (see Appendix B). As discussed earlier, 15 review pairs were presented to the participants sequentially in a random order. For each pair, the participants were asked to compare them and select the review that better matches their interest in the selected product feature. First, participants were asked a primary level question (see Figure 20) and provided the relative superiority of one review to another. The purpose of primary level question was to capture the overall preference

of the participants for a particular review in the review pair. The primary level question was asked as follows:

Q: Which review provides better information about the feature, you selected?

Upon answering the primary level question, they were asked three secondary level questions (i.e. the 3 Likert scale questions) and provided the relative superiority of one review to another with respect to three specific measures (i.e. relevance, knowledge and satisfaction). The questions were framed as follows:

(1) Review A provides more relevant information about the selected feature than review B

(2) Review A provides more knowledge about the selected feature than review B

(3) I am more satisfied by the information provided by review A than review B about the selected feature

A seven-point Likert scale from ‘‘strongly disagree’’ to ‘‘strongly agree’’ was used to capture the score given by each participant on the review pair (see Appendix B). The participants were also asked a question about the reviews, such as “In which review, the reviewer has provided negative comments on the selected product feature?” to make sure that they actually read the content of the reviews before they answered the primary and secondary level questions.

This user study has used 3 widely used system evaluation measures (i.e. relevance, knowledge, and satisfaction) to capture participants’ feedback on the quality of the

reviews. However, there are several other measures used in IS research for system evaluation and including those in this research would certainly enhance the quality of the evaluation process. But with the limited resources and other constraints such as a reasonable time frame, cognitive overload of the participants etc., the scope of the research has been restricted to the use of these three measures. In future, more resources will be allocated to include several other measures to conduct a comprehensive user study.

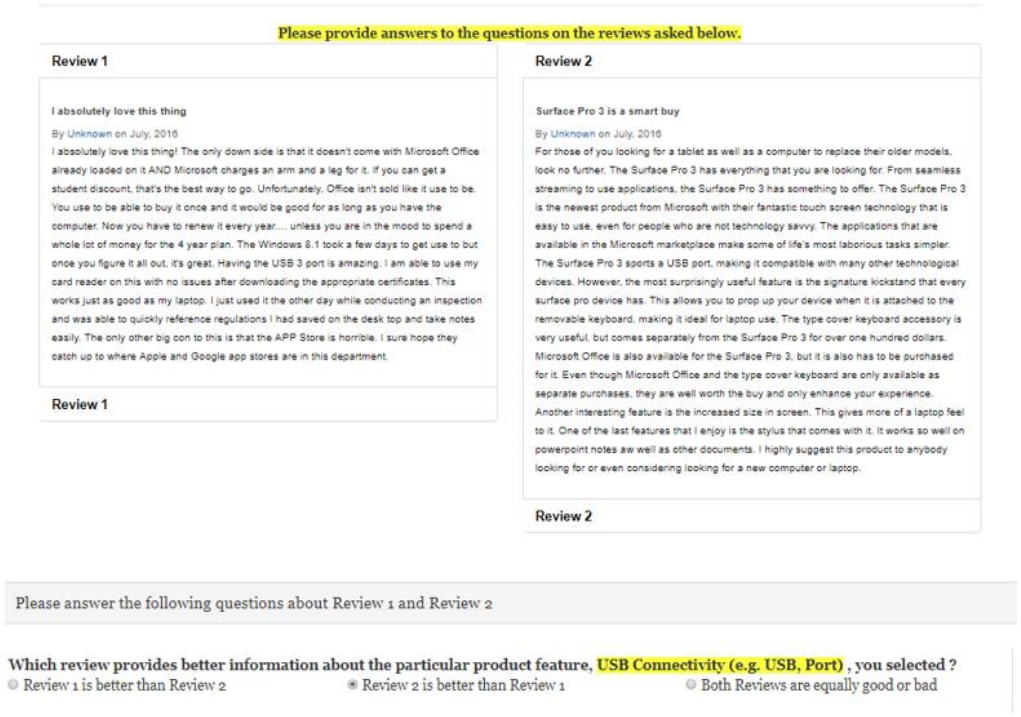


Figure 20. Primary level question

In addition to that, to avoid any kind of bias, review pairs shown to the participants were also randomized for each participant. An example “review preference” dataset after capturing the answers from a single participant is shown in Table 20.

Table 20. Basic review preference using pairwise comparison				
Comparison No.	Review a	Review B	Preferred Review	Scale (AHP)
1	R2	R1	R1	5
2	R3	R1	R1	4
….	….	….	….	…
15	Rn	R15	Rn	7

5.2.4 Determine the Preferred Ranking of the User

In this step, Analytical Hierarchical Processing (AHP) was applied to develop a rank of the 6 reviews for each participant (i.e. user ranking). AHP uses pairwise comparison method to measure the weight of two criteria by measuring the relative importance of one criteria over the other. In this case, AHP was applied to weigh reviews using the scales provided by the participants in the pairwise comparisons and assign a score to each review. The reviews were then ranked by these scores. The process is depicted by Figure 21. As AHP involves with the modeling of subjective judgements by the users, it is necessary to check the consistency in the subjective judgements.

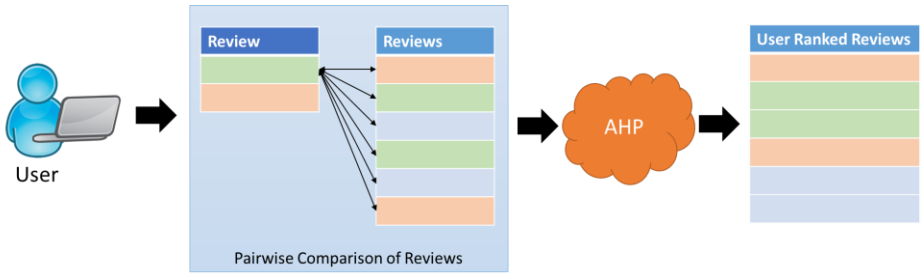


Figure 21. Getting reviews ranked by the user in a user study

AHP provides a measure called Consistency Ratio (CR) at each level of comparison that determines the level of consistency in terms of proportionality and transitivity in the subjective judgements by a decision maker [107] (i.e. participant). A low

consistency ratio indicates that the pairwise comparisons are adequately consistent, hence the subjective judgements of the participants would be acceptable.

Table 21. Priorities and ranking consistency using AHP		
Reviews	Weight from pairwise matrix eigenvalues	Participant's User Ranking
R1	0.2533351	1
R2	0.1266676	6
R3	0.1266676	4
R4	0.1436136	5
R5	0.1436136	3
R6	0.2061025	2
Consistency Ratio (CR) = 0.03949082		

In the user study, consistency ratio was calculated first, to ensure that the answers given to the pairwise comparisons by the participants were consistent. Any highly inconsistent answers (i.e. having consistency ratio more than 0.15) were excluded from further analysis. Table 21 shows an example of the user ranking with review priorities and the ranking consistency ratio calculated using AHP.

In this study, for each participant, the answers to the primary level questions were used to obtain a primary level user ranking. Similarly, the answers to the secondary level questions were used to develop three secondary level user rankings (i.e. user relevance ranking, user knowledge ranking and user satisfaction ranking). Table 22 shows an example analysis of secondary level user rankings and the associated consistency ratios.

Table 22. Calculating secondary level of rankings using AHP						
Reviews	Relevance priorities	Relevance ranking	Knowledge priorities	Knowledge ranking	Satisfaction priorities	Satisfaction ranking
R1	0.1644	2	0.0944	4	0.1644	2
R2	0.1469	6	0.2460	2	0.1469	6
R3	0.1469	5	0.1254	3	0.1469	5
R4	0.1644	3	0.0333	6	0.1644	3
R5	0.1644	4	0.0463	5	0.1644	4
R6	0.2127	1	0.4542	1	0.2127	1
Consistency Ratio (CR)		0.04		0.08		0.01

5.2.5 Grade the Review Ranking Systems

After obtaining consistent user ranking of each participant, next step was to find which one of the two system generated rankings ((i.e. helpfulness ranking and LCR ranking) ranked reviews in a more similar way to the user ranking. To achieve this, normalized kendall tau distance (*d*) was used to calculate the distance between user ranking and two system generated rankings as shown in Figure 18. Kendall tau distance (*d*) is a metric that counts the number of pairwise disagreements between two ranking lists after performing some bubble sorting swapping. The calculation for normalized kendall tau distance is given by:

$$d = \frac{n_d}{n(n-1)/2} \quad (1)$$

Where n_d the number of discordant pairs. For a pair (i, j) if observation i is ranked above observation, then the pair is called concordant. Otherwise, it is marked as discordant. Given this definition, if $d = 0$ both lists are perfect in agreement. If $d = 1$

one list has a complete reverse order than the other list. Thus, in this case, the system generated ranking which has less distance from the user ranking was considered as the preferred review ranking.

Table 23. Distance *(d)* between primary level user ranking and system generated rankings

	User Ranking	Review ID	Helpfulness Ranking	LCR Ranking
	4	R1	6	3
	2	R2	5	1
	3	R3	1	4
	6	R4	2	2
	5	R5	3	5
	1	R6	4	6
Distance (*d*)			-0.333	0.0666

In this user study, distance was computed both for primary level and secondary level user rankings. First, kendall tau distance (*d*) was computed between the primary level user ranking and the two system generated rankings for each participant as shown in Table 23.

Table 24. Distance (*d*) between user ranking and system generated rankings at primary level

	Participants	distance (d) with helpfulness ranking	distance (d) with LCR Ranking
	Participant 1	-0.466	-0.25
	Participant 2	0.122	-0.34
	Participant 3	-0.301	-0.178
	…..	……..	…….
	Participant n (n=66)	0.004	0.066

Next, distance was computed for the secondary levels user rankings using the same procedure. More specifically, distance was computed between, (i) user relevance

ranking and the system generated rankings, (ii) user knowledge rankings and system generated rankings, and (iii) user satisfaction ranking and system generated rankings.

5.3 Participants

A total of 66 participants (online panel of professional survey takers at Amazon Mturk) were recruited for the user study on a voluntary basis with an IRB approval.

Table 25. Participants background information

Total Participants = 66		Number of Participants
Age	< 18	0
	18-25	12
	26-35	24
	36-45	21
	46-55	8
	56-65	1
	>65	0
Education Level	Graduate	8
	Undergraduate	37
	High school/Diploma	18
	Other	3
Gender	Male	37
	Female	29
Familiarity with review websites	Moderately	9
	Strongly	18
	Extremely	39
Frequency browsing online product reviews	Daily	14
	Weekly	34
	Monthly	18
Impact of online reviews on product purchase	To a moderate extent	31
	To great Extent	19
	To very great extent	16

To ensure data quality, only those MTurk workers with at least 5,000 approved Human Intelligence Tasks (HIT) and a 97% of HIT approval rate were allowed to participate in the user study. The user study also made sure that any MTurk worker could not participate more than once in the user study. All the qualified MTurk

workers were further asked to fill out a pre-study questionnaire to ensure that all participants had prior experience with using online review platforms/online reviews and online shopping.

Participants received up to $10 in cash (through Amazon MTurk) after successful completion of the study. Among the participants 29 (i.e. 43 %) were female and 37 (i.e. 57 %) were male. Most of the participants were in the age range of 26-45 (i.e. 68 %). In general, almost 86 % of all participants were strongly familiar with the review websites and 51 % of participants browsed reviews about a product before making purchase decisions on a weekly basis. Also, for 53 % of the participants, online reviews had a strong impact on their product purchase decisions. Table 25 summarizes the background information of all participants in the study.

5.4 **Experimental Procedure**

This user study was a web-based user study where participants conducted the user study online. They were instructed to read through each steps of the user study thoroughly before proceeding with the experiment. Participants were requested to complete the user study in one sitting and were able to ask us any questions through email at any time during the study.

1. First all participants browsed “www.curetext.com” website and were asked to answer a few questions to determine their eligibility to participate in the user study. In particular, participants filled in a pre-questionnaire on their demographic information (including age and gender) and experience with

review platforms such as amazon.com to buy products (see Appendix B). Participants not eligible for the study, were politely requested to withdraw from the study.

2. Eligible participants, (Amazon MTurker) were auto registered based on their worker id before they start the user study. Then participants were asked to read a disclaimer suggesting that they had to follow all the steps of the user study correctly and provide answers to the questions asked to them during different steps of the user study correctly. To ensure that participants were answering to the questions correctly, a method (AHP consistency ratio) was employed that checked consistency in their answers. Participants who provided highly inconsistent answers, (incorrect answers) were not compensated and their feedback was not incorporated in the analysis (see Appendix B).

3. On agreement with the disclaimer, brief objective/aim of the study was presented to the participants first (see Appendix B). Next, online consent was obtained from each participant who wanted to participate in the study. Participants were allowed to withdraw from the study at any time and the data entered by each participant were removed and excluded for any kind of further analysis. After providing their consent to continue with the study, participants were shown the procedure/tasks of the experiment followed by the demo of the experiment. A web portal was provided to the participants with a training/demo session of how to select product features, compare reviews and

provide their answers to the questions regarding the reviews. Participants were able to view the demo whenever they needed throughout the experiment by visiting the demo page at http://www.curetext.com/home/showdemo. After viewing the demo, participants went through a pre- experiment test, where they were asked two questions regarding the demo video to ensure that the participants correctly understood the tasks. They were able to continue with the experiment only after providing correct answers to the questions. Also, participants were instructed to ask any questions through email if they had any problem with understanding the tasks/procedure of the study before they started the experiment.

4. Upon successfully providing correct answers, participants were directed to a page that shows a product name. Participants were then requested to select the product to see some description, technical specifications, visual of the product and a list of features of the product (the information of the product was exactly the same as it was posted in the original Amazon.com website). After instructed to read the information about the product, participants were asked to choose only one of the product listed features. On selecting a particular feature that they liked, they were presented with a set of review pairs (15 pairs of reviews) sequentially (see Appendix B).

5. Participants were asked to thoroughly read a pair of reviews and provide their answers to the questions related to the presented reviews. Also, an additional question was asked about the two reviews in each review comparison process

to ensure that the participants are actually reading the reviews before they provide their answers to the questions. Once the participant completed the review comparison for the first product, the same steps were repeated for the second product.

After the experiment, the Amazon Mturk participants got compensated by typing a unique code provided in the user study web portal.

Chapter 6. Data Analysis and Results

This chapter presents data analysis and results of the user study. The data collected in the user study were statistically analyzed by AHP and normalized kendall tau distance measure to determine the efficacy of FDPPR. This chapter is divided into three subsections as follows. Sections 6.1 analyzes the data from the participants to determine whether the answers provided by them were consistent. Section 0 describes the results of primary level user ranking analysis. In section 6.3, the results of the secondary level user ranking analysis are presented. The secondary level user ranking analysis was the basis for testing the three hypotheses proposed in the user study.

6.1 Results on Consistency

As discussed in Section 5.2.4 in pairwise comparisons, inconsistent answers (i.e. a consistency ratio more than 0.15) were removed before further data analysis. In the primary level ranking, 59 participants for product 1 and 62 participants for product 2 provided consistence answers (see Table 26). Thus, primary level user rankings corresponding to participants who provided consistent answers were only taken for further analysis.

Table 26. Participants' consistency ratios (CR) for primary level user ranking

	Total Participants	Participants with CR < 0.15	Percentage	Minimum CR	Maximum CR
Product 1	66	59	89%	0	0.14
Product 2	66	62	94%	0	0.13

In the secondary level user rankings, consistency ratios were only calculated for the participants who already provided consistent answers in the primary level ranking. In

the secondary level user rankings, about 66% of participants in both products provided consistence answers for relevance. For knowledge, while 54% provided consistence answers in product 1, about 48% of participants provided consistent answers for product 2. For satisfaction, 42% provided consistence answers for product 1 and 62% provided consistence answers for product 2. Table 27 also shows the range in the consistency ratios for secondary level of rankings. From both levels of user rankings, the result showed that some participants provided perfectly consistent answers (consistency ratio of 0) while other participants provided consistent answers at a permissible level (i.e. CR less than 0.15).

Table 27. Participants' consistency ratios for secondary level user ranking

	Type of Ranking	Number of consistent participants (%)	Consistency Ratio (CR)
Product 1	Relevance	39 (66.1%)	Min CR = 0.0 Max CR =0.14
	Knowledge	25 (42.3%)	Min CR = 0.0 Max CR = 0.14
	Satisfaction	32(54.2%)	Min CR =0.0 Max CR =0.14
Product 2	Relevance	41 (66.1%)	Min CR = 0 Max CR = 0.14
	Knowledge	30 (48.38%)	Min CR = 0.0 Max CR = 0.13
	Satisfaction	39 (62.09%)	Min CR = 0 Max CR = 0.14

6.2 Results on Primary Level User Ranking

The primary level user ranking data were analyzed to find out the overall performance (i.e. ranking reviews in a more similar way to a participant's self-ranking) of the system generated rankings. As discussed in section 2.3.2, for each participant, normalized kendall tau distance (d) was calculated between user ranking

and the system generated rankings. The system generated ranking having less distance from user ranking was considered as the better and desired ranking for the participants. In this analysis, first distance (d_{help}) was calculated between user ranking and helpfulness ranking. Next, distance (d_{LCR}) was computed between user ranking and LCR ranking for each participant (see Appendix C and Appendix D). A paired sample t-test (see Appendix E) was performed on this data that showed that average distance between LCR ranking and user ranking was significantly less than the average distance between helpfulness ranking and user ranking. For product 1, $t_{58} = 5.2$ ($p < 0.05$) and product 2, $t_{61} = 6.9$ ($p < 0.05$). This result implied that, for each product, LCR ranking had less distance from user ranking than helpfulness ranking. The data analysis on primary level user rankings also showed that more than 70% of user rankings were close to the LCR ranking in product 1. A similar result was also obtained for product 2, where more than 80% of user rankings were close to the LCR ranking. From the analysis, it was confirmed that LCR ranking was close to participants' self-rankings (i.e. user rankings). Thus, it can be concluded that overall participants preferred the proposed LCR ranking than the existing helpfulness ranking.

6.3 Hypotheses Test Results

The secondary level user rankings data were analyzed to test the three hypotheses proposed in the user study (see Section 5.1). The hypotheses were tested to understand the participants' preference for a particular system generated ranking with respect to each of the three measures (i.e. relevance, knowledge, and satisfaction).

The same data analysis procedure as presented in Section 6.2 was performed to test the 3 hypotheses.

In the next section, the details of the hypotheses tests are presented. Table 28 presents the results of the hypotheses test.

Table 28. Hypotheses test results (results on participants' secondary level user rankings)

	Hypotheses	Measures	Expected Outcome	Result of t-test	Supported?
Product 1	H1: Consumers will be able to access more relevant reviews from the top reviews ranked by the personalized review ranking (i.e., LCR ranking) as compared to the top reviews based on helpfulness ranking	Relevance	Distance between (LCR ranking and user ranking) < Distance between helpfulness ranking and user ranking)	Significant (< 0.05)	Y
	H2: The top reviews ranked by personalized review ranking (i.e. LCR ranking) will provide a higher level of knowledge about interested features of a product to consumers than those ranked by helpfulness ranking	Knowledge	Distance between (LCR ranking and user ranking) < Distance between helpfulness ranking and user ranking)	Significant (< 0.05)	Y
	H3: Personalized review ranking (i.e. LCR ranking) will lead to higher consumer satisfaction than helpfulness ranking	Satisfaction	Distance between (LCR ranking and user ranking) < Distance between helpfulness ranking and user ranking)	Significant (< 0.05)	Y

Table 28. Hypotheses test results (results on participants' secondary level user rankings)					
Product 2	H1: Consumers will be able to access more relevant reviews from the top reviews ranked by the personalized review ranking (i.e., LCR ranking) as compared to the top reviews based on helpfulness ranking	Relevance	Distance between (LCR ranking and user ranking) < Distance between helpfulness ranking and user ranking)	Significant (< 0.05)	Y
	H2: The top reviews ranked by personalized review ranking (i.e. LCR ranking) will provide a higher level of knowledge about interested features of a product to consumers than those ranked by helpfulness ranking	Knowledge	Distance between (LCR ranking and user ranking) < Distance between helpfulness ranking and user ranking)	Significant (< 0.05)	Y
	H3: Personalized review ranking (i.e. LCR ranking) will lead to higher consumer satisfaction than helpfulness ranking	Satisfaction	Distance between (LCR ranking and user ranking) < Distance between helpfulness ranking and user ranking)	Significant (< 0.05)	Y

6.3.1 Results of H1: Personalized Review Ranking vs. Traditional Helpfulness Ranking with respect to Relevance

H1 posits that consumers will be able to access more relevant reviews from the top reviews ranked by the personalized review ranking (i.e., LCR ranking) as compared to the top reviews based on helpfulness ranking. To test the hypothesis, normalized kendall tau distance (*d*) was calculated between user relevance ranking and the two

system generated rankings (i.e., helpfulness ranking and LCR ranking) for each participant. The system generated ranking having less distance from user relevance ranking was considered as the better and desired review ranking. First, distance (d_{help_rel}) was computed between user relevance ranking and helpfulness ranking. Next, distance (d_{LCR_rel}) was computed between the user relevance ranking and LCR ranking for each participant (see Appendix F and Appendix I). A paired sample t-test was performed on this data that confirmed that the average distance between the LCR ranking and user relevance ranking was significantly less that the average distance between helpfulness ranking and user relevance ranking (see Appendix E). For product1, $t_{36} = 2.2$ and $p < 0.05$; and for product2, $t_{40} = 4.6$ and $p < 0.05$). So, the result of the analysis indicated that compared to helpfulness ranking, LCR ranking ranked reviews by their relevance in a more similar way to a participants' self-ranking of the same reviews (i.e. user relevance ranking). Thus, it was concluded that participants obtained more relevant reviews by using the LCR ranking than helpfulness ranking; thus, hypothesis 1 was supported in this research.

6.3.2 Results of H2: Personalized Review Ranking vs. Traditional Helpfulness Ranking with respect to Knowledge

H2 posits that top reviews ranked by personalized review ranking (i.e. LCR ranking) will provide a higher level of knowledge about interested features of a product to consumers than those ranked by helpfulness ranking. For testing the hypothesis, the same procedure was repeated as in H1 (see Section 6.3.1). Normalized kendall tau distance (d) was computed between user knowledge ranking and the system generated rankings for each participant. First, distance (d_{help_know}) was computed between user

knowledge ranking and helpfulness ranking followed by the distance (d_{LCR_know}) calculation between user knowledge ranking and LCR ranking for each participant (see Appendix G and Appendix J). A paired sample t-test (see Appendix E) was performed that showed that the average distance between LCR ranking and user knowledge ranking was significantly less than the average distance between helpfulness ranking and user knowledge ranking. The results of the t-tests were highly significant (i.e. for product 1, $t_{24} = 1.88$ ($p < 0.05$) and for product2, $t_{29} = 4.23$, ($p < 0.05$)) for both of the products. So, the result of the analysis indicated that LCR ranking ranked reviews by their knowledge in a more similar way to a participant's self-ranking of the reviews (i.e. user knowledge rankings). Thus, it was concluded that LCR ranking was more effective in showing reviews to the participants that had more knowledge about the product features selected by them than helpfulness ranking. Hence, H2 was also supported in this research.

6.3.3 Results of H3: Personalized Review Ranking vs. Traditional Helpfulness Ranking with respect to Satisfaction

H3 posits that personalized review ranking (i.e. LCR ranking) will lead to higher consumer satisfaction than helpfulness ranking. To test the hypothesis, normalized kendall tau distance (*d*) was again computed between user satisfaction ranking and the two system generated rankings for every participant (see Appendix H and Appendix K). In line with the results of H1 and H2, the result of the analysis showed that, for both product 1 and product 2, LCR ranking was closer to the user satisfaction ranking than helpfulness ranking. The result of the paired sample t-test (see Appendix E) confirmed that the average distance between LCR ranking and user satisfaction

ranking was significantly less that the average distance between helpfulness ranking and user satisfaction ranking. For product 1, $t_{31} = 1.81$ ($p < 0.05$) and for product 2, $t_{38} = 3.06$ ($p < 0.05$). The result of the analysis indicated that LCR ranking ranked reviews by their satisfaction level in a more similar way to a participants' self-ranking of the same reviews (i.e. user satisfaction rankings). Hence, the result of the analysis implied that LCR ranking provided more satisfactory reviews to the participants than helpfulness ranking, based on their selected product features. Thus, hypothesis 3 (H3) was also supported in this research.

6.4 Summary of the Study Results

From the above data analyses, it was found that personalized review ranking (i.e. the LCR ranking) was significantly closer to user rankings (i.e. participant's self-ranking of the reviews) than helpfulness ranking. First, the result of the primary level user ranking analysis indicated that overall, LCR ranking proved to be a better and desired ranking by the participants compared to helpfulness ranking. Similarly, the results of the secondary level user ranking analyses (see Table 28) showed that LCR ranking performed significantly better than helpfulness ranking. Based on these results, it was concluded that LCR ranking provided more relevant, knowledge rich and satisfactory reviews to the participants than helpfulness ranking.

Chapter 7. Discussion

This research thoroughly examined and explained the need for a personalized review ranking system that provides more relevant and helpful reviews to the consumers based on their personal preferences. Towards this end, FDPPR, a personalized review ranking framework was constructed and applied to several product review data. This research also evaluated the effectiveness of the personalized review ranking framework by conducting a user study that compared the performance of the personalized review ranking with respect to the most widely used helpfulness ranking (i.e., people found useful). To the best of the author's knowledge, this is the first study that provides a personalized framework for predicting the helpfulness of consumer reviews based on product features of a product. In addition, by conducting the user study on the real product review data, it is also set as the first empirical study that examined the possible benefits consumers get by using a personalized review ranking system. This chapter summarizes the user study results, then discusses the implications for theory and practice, along with its limitations and future research possibilities.

7.1 Summary of Findings

The overall findings of the study showed that, personalized review ranking (i.e. LCR ranking) was significantly closer to user ranking (i.e. each participant's own ranking of reviews) than helpfulness ranking for both products.

First, the results of the user study showed that, when providing relevant reviews to the participants based on their product feature preference, personalized review ranking (i.e. LCR ranking) was significantly close to user ranking than the traditional helpfulness ranking. Specifically, the results supported the proposed hypothesis (H1) that posits that consumers will be able to access more relevant reviews from the top reviews ranked by the personalized review ranking as compared to the top reviews based on helpfulness ranking.

Second, a significant evidence was also found to support the second hypothesis (H2) that posits that the top reviews ranked by the personalized review ranking provide a higher level of knowledge about interested features of a product to consumers than those ranked by helpfulness ranking. As shown by the user study results, personalized review ranking was also significantly close to user ranking than the traditional helpfulness ranking.

Third, the results of the study also showed that personalized review ranking was significantly close to user ranking than the traditional helpfulness ranking while providing satisfactory reviews to the participants based on their product feature preferences. This indicated that personalized review ranking provided higher consumer satisfaction than helpfulness ranking, thereby supported the third hypothesis (H3) of this research.

7.2 Theoretical Implications

This research makes three main theoretical contributions. First, this study offered a new lens to examine review helpfulness. Compared with most prior studies, which focused on exploring a series of review characteristics influencing review helpfulness [15, 17, 24], this study, as one of the first attempts, theorized and tested the combined effect of product features and review characteristics on the helpfulness of the reviews. The user study showed that both product features and review characteristics of reviews were important when predicting readers' perceptions of the helpfulness of reviews.

Second, this study is the first to predict helpfulness of online reviews in a personalized way using a latent class modeling approach. This study extended the earlier studies [11, 24] with its application of a latent class regression model for verifying the combined effects of product features and review textual characteristics on helpfulness of the reviews. LCR analysis used in this research, provided insights to wade through large volume of online reviews and extract only the relevant and helpful reviews that cater to the consumer's specific information needs. The results of this study confirmed and extended the knowledge about the varied impact of the review characteristics on the helpfulness of the reviews for different consumer classes. Moreover, this research provided an effective way to reduce the difficulty associated with the great variation in the review content and review quality.

Third, this research took a user driven system evaluation approach to test the real efficacy of the proposed framework of review ranking personalization. Despite

expensive and time consuming, user driven studies have great potential as end users get directly involved with the system. The effectiveness of the system is validated by the users' personal satisfaction with the system. Most helpfulness prediction models in the literature used a model based evaluation strategy to evaluate the performance of the constructed models. This research is the first to conduct a comprehensive user study to capture users' perceptions towards the personalized review ranking system that was designed to provide more relevant, satisfactory, and knowledge-rich reviews to the users based on their specific needs (i.e. product feature preferences). The user study followed an evaluation framework where the performance of the personalized review ranking system was captured by using widely-accepted evaluation measures, such as relevance, knowledge and satisfaction and analyzed by using well-known statistical methods.

7.3 Practical Implications

This research will help multiple stakeholders of ORPs (e.g. perspective consumers, retailors and the manufactures of the product) in searching, organizing, and presenting product reviews effectively.

The research will help prospective consumers get personalized ranking of reviews that better suit their product feature preferences when looking for new products to purchase. This will lower the information overload and help consumers make better and quicker purchase decisions. The research will facilitate the design of user interface by which prospective consumers will be able to specify their product feature preferences when searching reviews for a product. The research will provide greater

flexibility to the prospective consumers by allowing them to change their product features preferences dynamically and see a different customized review list based on their changed product feature preferences.

From the retailer's perspective, the research will assist retailors of the system (e.g., Amazon.com) in better organizing and presenting consumer reviews to prospective consumers. Several product features of the product will be automatically extracted and identified from consumer reviews that improve the understanding of the product by the retailors from the consumers' point of view. Organizing consumer reviews around the identified product features will enable a feature-based review search that will help individual prospective consumers find their personalized information easily and decrease the cost of information search. The retailors can also track the preferred product features of a consumer and can provide better product recommendations to individual consumers.

From the manufacturers' perspective, the research will help them better understand consumers' needs. Companies will also find out the product features most discussed by the reviewers, which will help businesses improve product features in the next release of the product. This research employed techniques to identify various features and find consumers' opinions (positive or negative) concerning the features. Analyzing the negative opinions on the product features will enable manufacturers to understand consumers' needs and release improved products in the future that will increase sales.

7.4 Limitations and Future Work

Due to limited resources and other constraints, this research has also several limitations that serves as suggestions for future work. First, in this research, product technical features were only extracted and used in the FDPPR modeling framework to identify consumer classes in the reviews. Usually text reviews contain many non-technical features such as “small” for size of a product. Considering all possible kind of features in the LCR modeling process could enhance the results of the analysis in terms of finding more meaningful consumer segmentation and their associated characteristics. This research involved with extracting product features from review text using a graphical model, LDA. In future, other approaches such as simple term frequency–inverse document frequency (tf/idf) or non-negative matrix factorization (NMF) techniques could be applied to find product features from the review text. In addition to that, in this research sentiments of the reviews were extracted at the document level (i.e. review level). In future, a fine-grained product feature level sentiment analysis could be carried out to better determine the sentiments of the reviews. Previous studies on review helpfulness prediction used many other review characteristics to predict the helpfulness of the reviews. However, in this research, only a subset of several well-established review characteristics was considered that contributed towards the helpfulness of the reviews in the review personalization framework. In future, the proposed FDPPR framework will be made more robust by considering several other review characteristics (e.g., number of features mentioned in review, number of times a feature has been discussed in a particular review, other lexical features such as number of adjectives/nouns/pronouns, number of spelling

mistakes present in a review etc.) as well as the reviewer's characteristics (see Section 2..1.1) while predicting the helpfulness of the reviews for different consumer classes. Also, post processing can be introduced to boost LCR ranking of certain reviews based on subjective rules derived from subject matter experts. In addition to that in future diversity based ranking techniques such as maximal marginal relevance (MMR) can also be applied to the LCR based ranking to maximize the relevance and novelty in the finally retrieved top-ranked reviews.

Although this research used a user centric approach to evaluate the performance of the FDPPR framework, it suffers from several problems associated with user centric system evaluations such as higher cost, longer time, and scalability. Thus, in future a hybrid approach could be applied to integrate a system centric approach with the current user-centric approach to check the real efficacy of the framework. In the user study, to minimize cognitive overload of the participants while reading long textual reviews and answering multiple questions during the study, only a small set of reviews were chosen to be ranked by the participant. In future, with the availability of enough resources and other methods (e.g. presenting succinct version of a review text without losing the context of the review), a large number of reviews could be selected for the evaluation of the framework. In addition, in this research, three measures (i.e. relevance, satisfaction, and knowledge) were used to capture the user feedback about the quality of the reviews. As a part of the future work, more objective/subjective measures could be included to in the user study to capture user feedback at the lower level of granularity that could potentially enhance the quality of the evaluation process.

Chapter 8. Conclusion

Frequently visiting review websites (ORPs), such as Amazon.com, has become a very regular habit of consumers while doing online/offline shopping. Online reviews provided by ORPs have become potential decision-making tools for consumers. Online reviews provide abundant information about products or services that consumers want to buy. However, the number of reviews in the review portals is increasing at a very rapid pace and creating a problem of information overload for consumers. Consumers are vigorously getting confused with a lot of information of varying quality and content while trying to assess the real value of a product. Consumers often evaluate the quality of a product based on specific product features of a product. These phenomena underline the need for developing a personalized review-ranking system that can provide relevant and helpful reviews to the users based on their personal preferences on specific product feature(s) of a product.

This research made several contributions to the field of personalized helpfulness prediction of online reviews and user-centered evaluation of proposed personalized helpful prediction model. First, a generic framework, FDPPR, for personalizing the ranking of online reviews was developed using a systematic approach. This framework provided a theoretical interpretation and a mathematical estimation technique to model the helpfulness of online reviews in a personalized way. Several NLP and text-mining approaches such as sentiment analysis, LDA were performed on the review data to quantify the reviews in terms of several product features and review textual characteristics. A statistically robust method, LCR, was utilized to

model the helpfulness of the review that could simultaneously use the extracted product features and the review textual data to predict helpfulness of the reviews based on the product feature preferences of the users. Second, the FDPPR framework was applied on several product review data to generate personalized review rankings for those products. Third, to understand the importance and benefits of developing a personalized review-ranking system, an empirical user study was conducted. The findings from this study suggested that review ranking personalization is an effective process of reducing information overload by providing relevant and helpful reviews to the consumers based on their product feature preferences. In addition, this study provided new insights about how to design and conduct a user study in the domain of review helpfulness prediction.

Appendix A. User Study Setup in Amazon mTurk

Amazon mTurk was used as the primary platform to recruit the participants for the study. For recruiting participants for the study first a "requester" account was created in the Amazon mTurk website (https://requester.mturk.com).

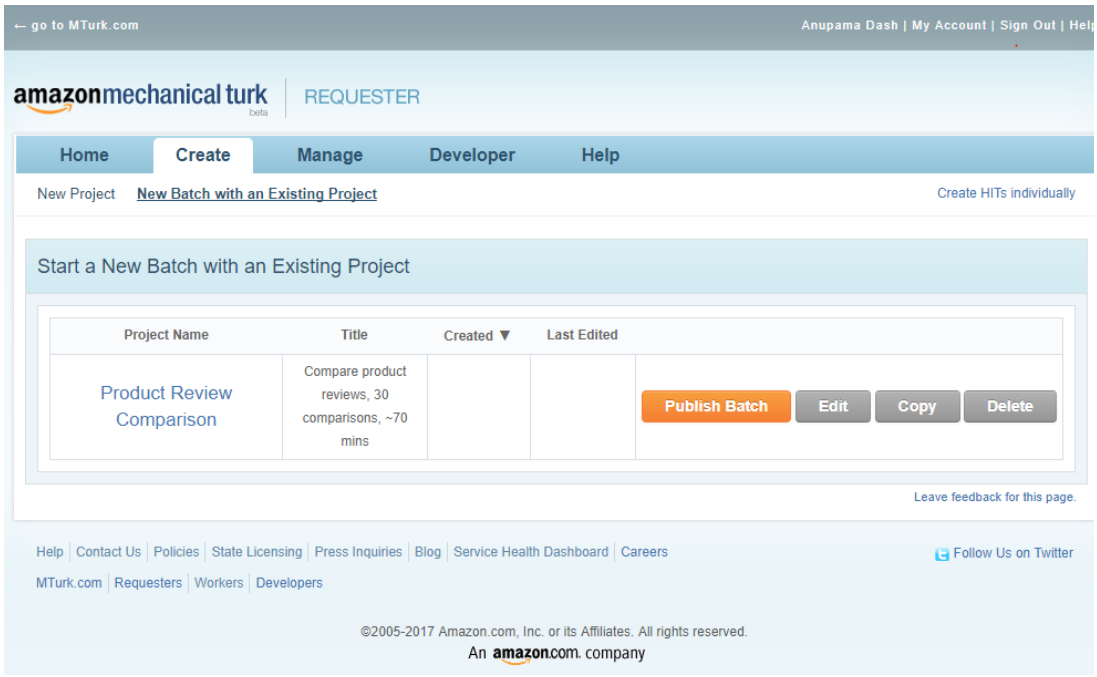


Figure 22. Creating a User Study Project in Amazon mTurk

Next a project was created to recruit the participants (see Figure 22) for the user study (i.e. a Human Intelligence Task or HIT). When setting up the study the reward for the HIT was setup.

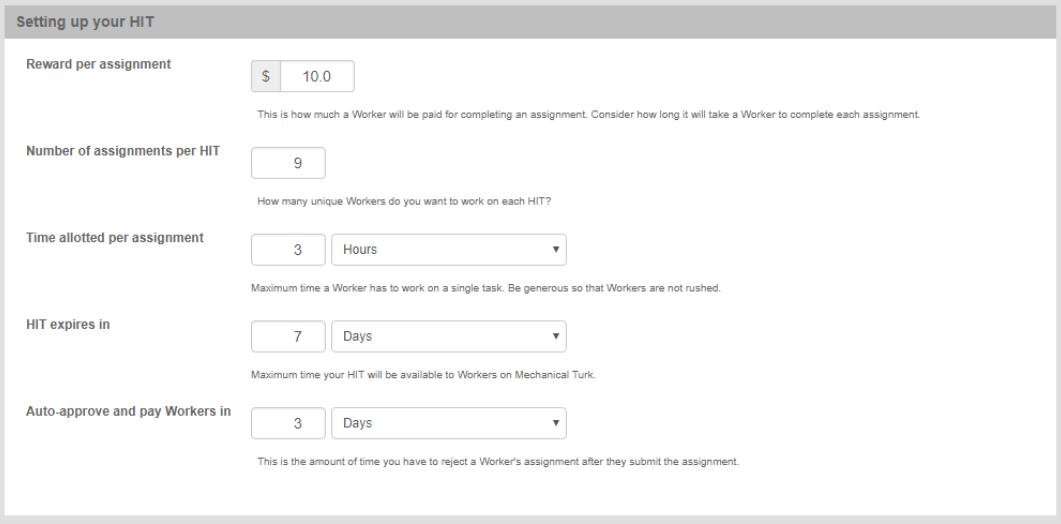


Figure 23. Setting Amazon mTurk HIT rewards for the user study

Figure 23 shows the setup of a HIT for the user study project created in Amazon mTurk. When setting up the HIT the worker requirements were set to an acceptable level to make sure only experienced good mTurk workers participate in the user study.

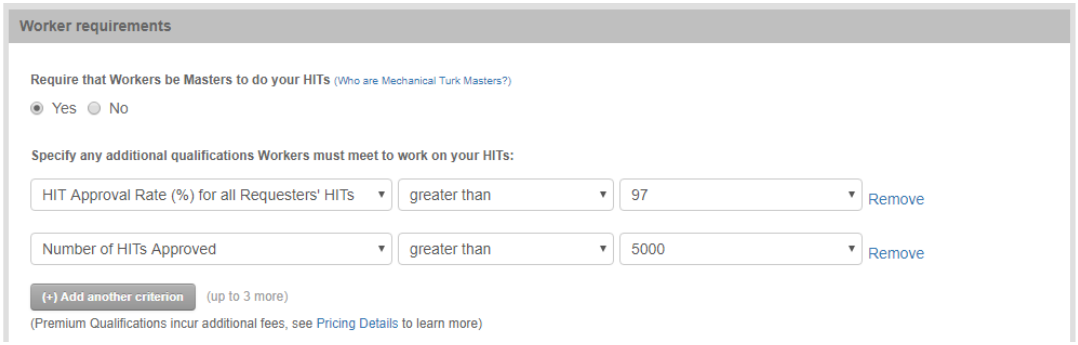


Figure 24. Amazon mTurk HIT worker requirements

In this user study the HIT approval rate for the Amazon mTurk worker was set at greater than 97 and the number of HITs approved was set at greater than 5000 (see Figure 24.

The design for the HIT contains (i) brief description for the Amazon mTurk worker about the user study and (ii) a Java Script code to get the mTurk works unique mTurk

id that needs to be sent to the user study website at http://www.curetext.com. The Amazon workers mTurk worker id is saved on the user study website to prevent repeat work by the workers.

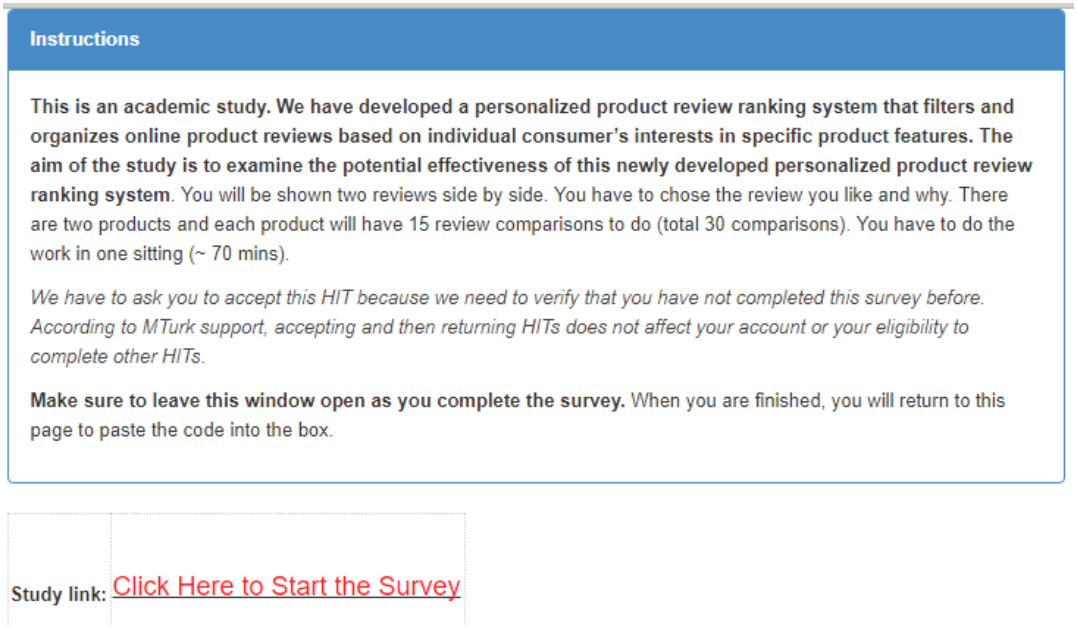


Figure 25. Instructions for the Amazon mTurk Worker

Figure 25 shows the instructions for the workers before they start the study. When the worker clicks on the survey link the Turk worker is taken to a ubique URL based on the worker's id, http://www.curetext.com/mturk/mturk-worker-id. The worker is them asked a set of eligibility questions to find out the worker's suitability for the user study.

Appendix B. User Study Website Details

The user study was conducted online at http://www.curetext.com, a website developed as part of this research, that (i) allows participants from Amazon mTurk as well as regular users to participate in the study, (ii) verify the eligibility of the users for the user study, (iii) get necessary consent from the users, and (iv) perform a comparison of reviews that span across two products, namely, Microsoft Surface Pro3 and Celestron SkyMaster Binoculars.

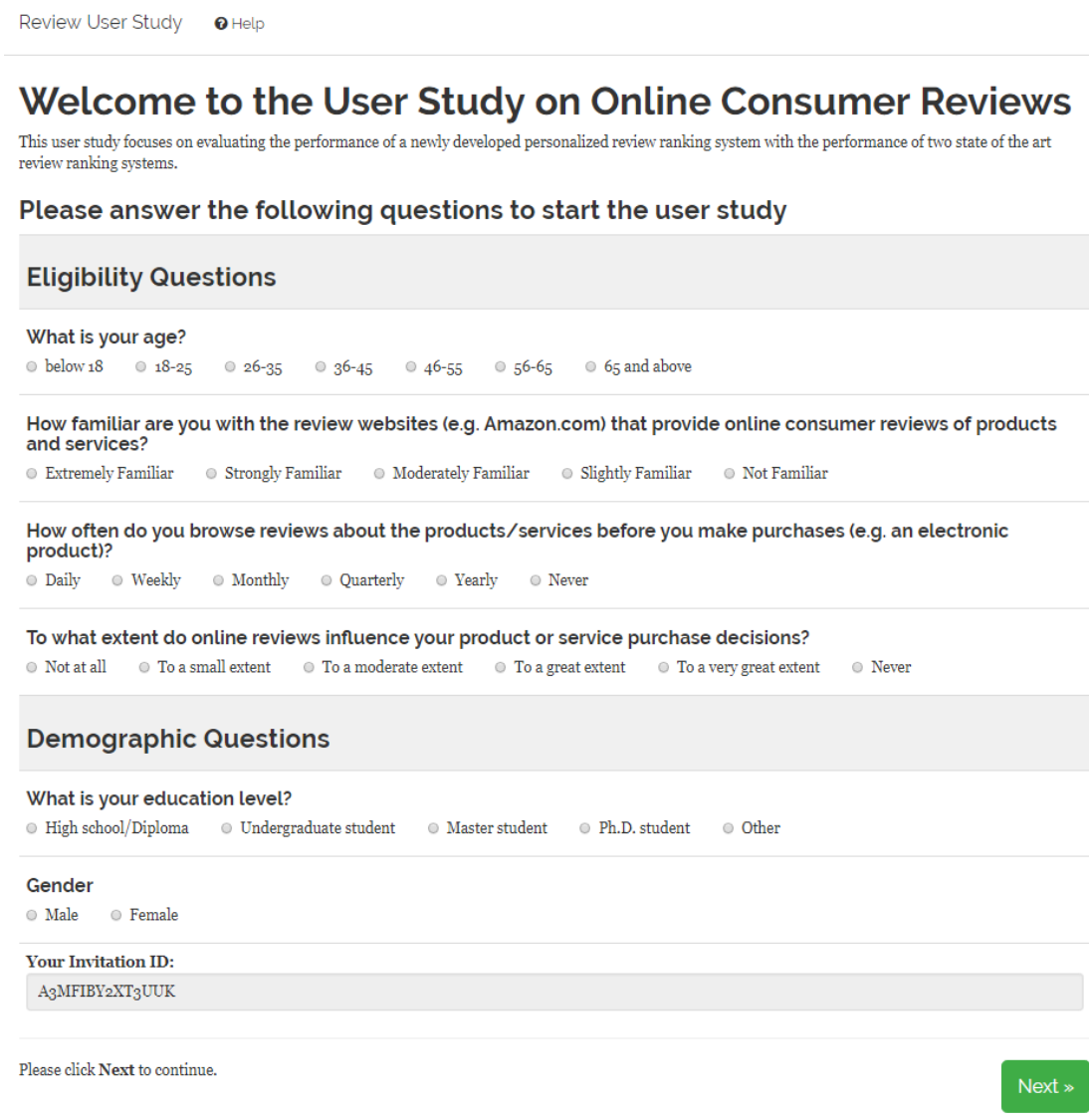


Figure 26. Eligibility questions for the participants

Figure 26 shows the eligibility questionnaires that are asked to every participant to determine their eligibility for the user study.

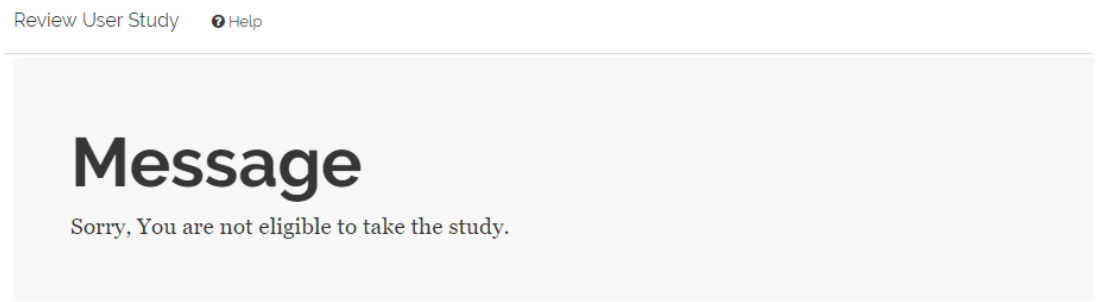


Figure 27. Message for the participants who are not eligible for the user study

If the participant is not eligible for the study he/she is shown a message notifying about it (see Figure 27).

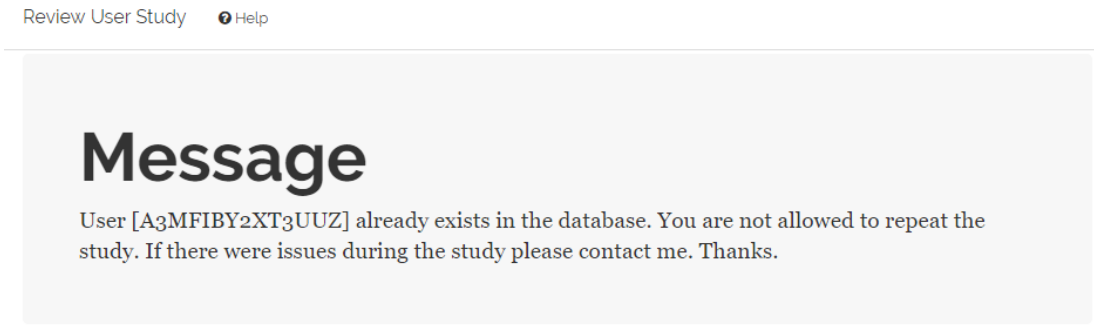


Figure 28. Message for the participants who try to repeat the user study

The user study can be done by a user only once. If the same user tries to do the user study again he/she is conveyed the message that it is not allowed (see Figure 28).

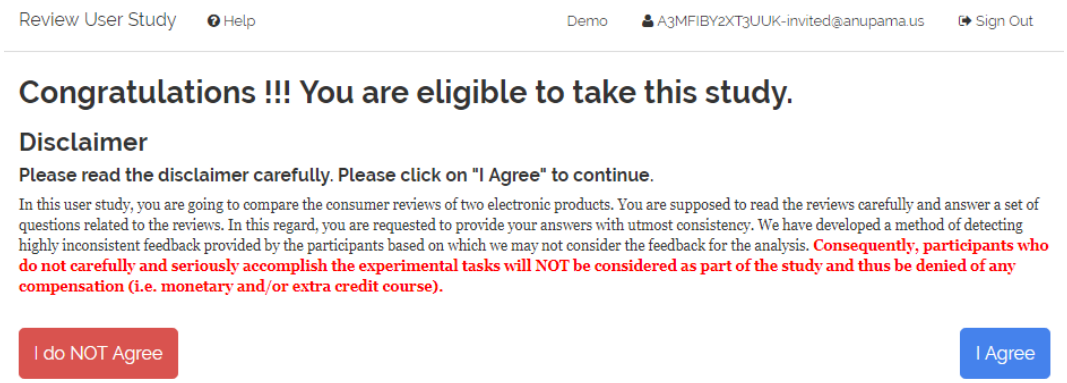


Figure 29. Disclaimer for the participants who are eligible for the user study

Once the participant is determined to be eligible for the user study he/she is shown the disclaimer and he/she has to agree to the disclaimer to continue to the study (see Figure 29).

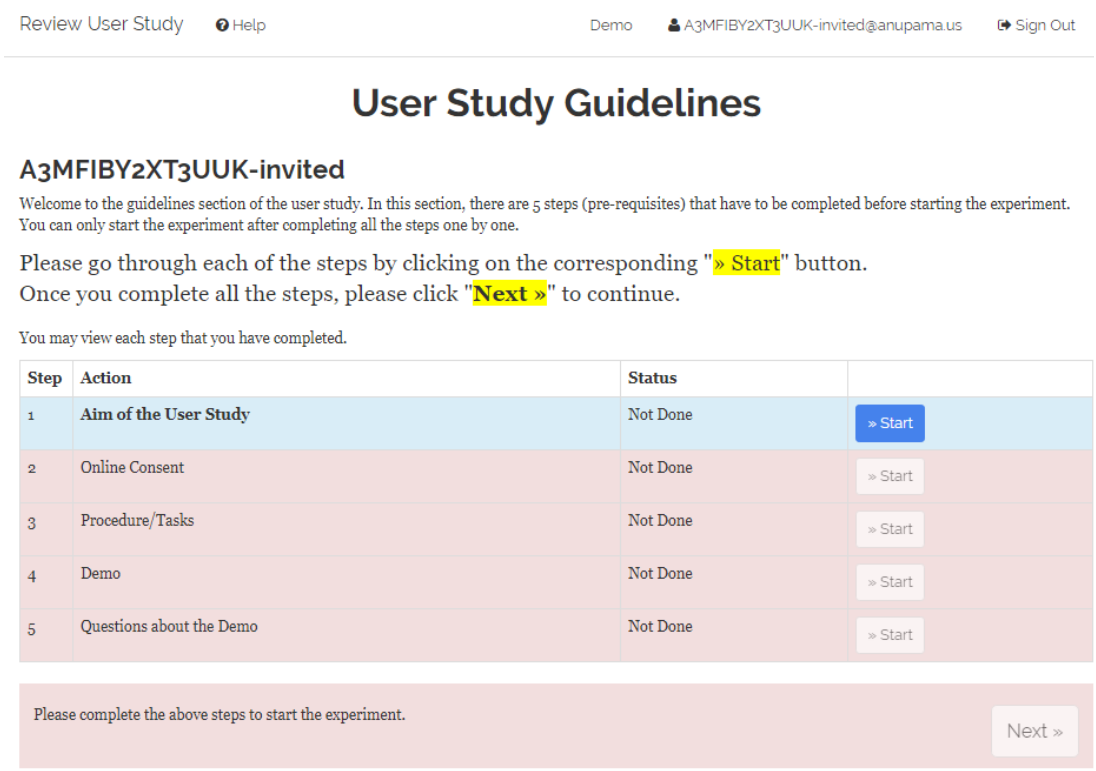


Figure 30. User study guidelines for eligible participants

Once the participant accepts the user study disclaimer he/she is shown the user study guidelines (see Figure 30). There are five steps in the user study guidelines, namely (1) Aim of the user study, (2) online consent, (3) procedure/tasks, (4) a video demo of the user study, and (5) some questions to the participants to make sure he/she understands how to do the user study.

The aim of the study is shown to the participants first (see Figure 31).

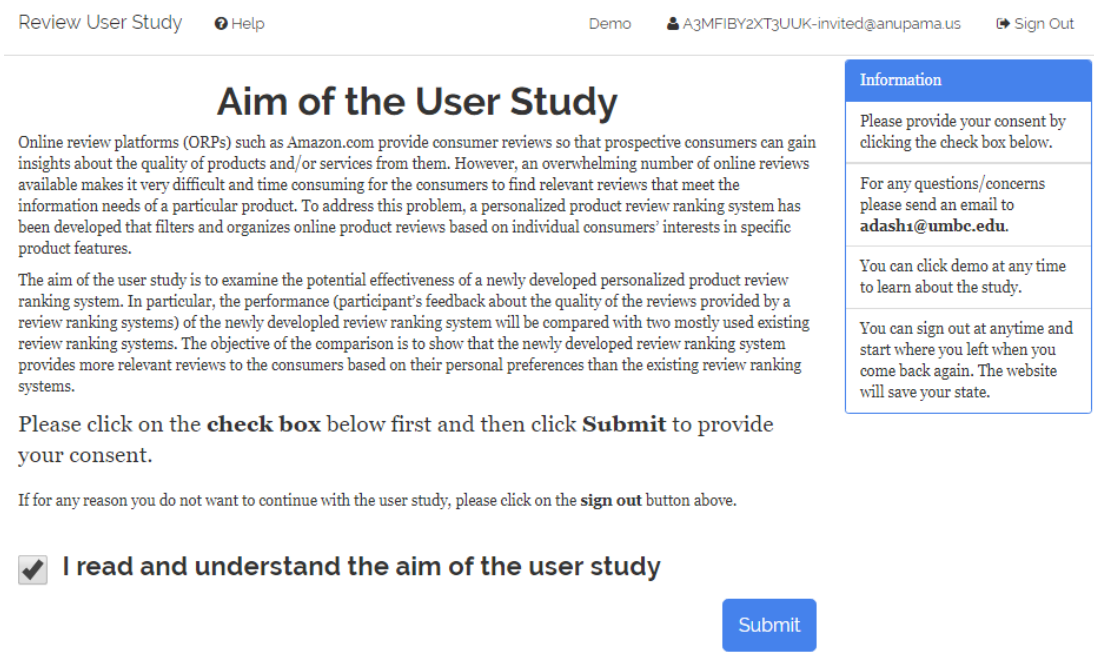


Figure 31. Aim of the user study

Once the participant reads and understands the aim of the user study he/she is taken back to the user study guideline page (see Figure 32).

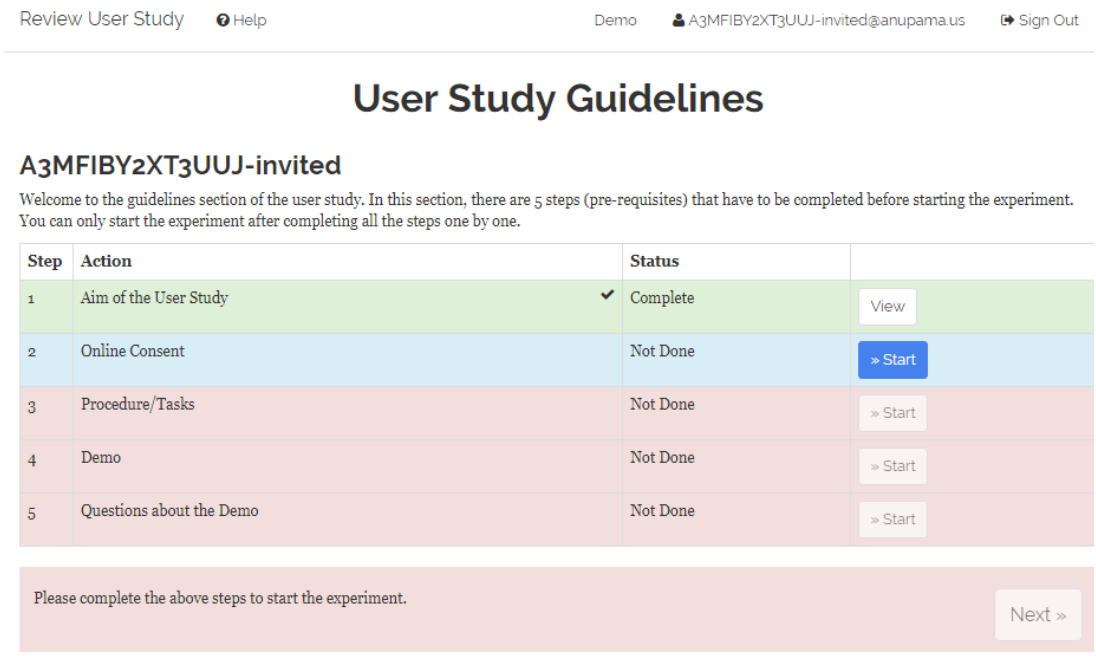


Figure 32. Completed aim of the user study

Next the participant is shown the "Online Consent" form (see Figure 33).

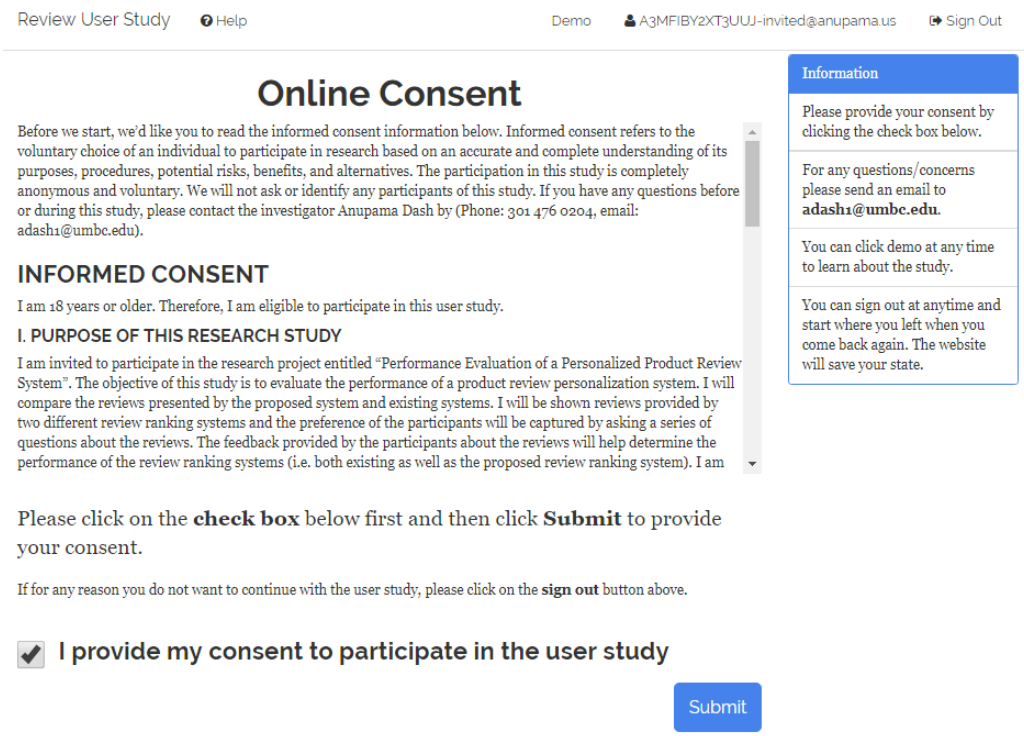


Figure 33. User study online consent form

Once the participant accepts the online consent form he/she is returned back again to the user study guidelines page (see Figure 34) which now shows that both (i) aim of the user study, and (ii) online consent is already done.

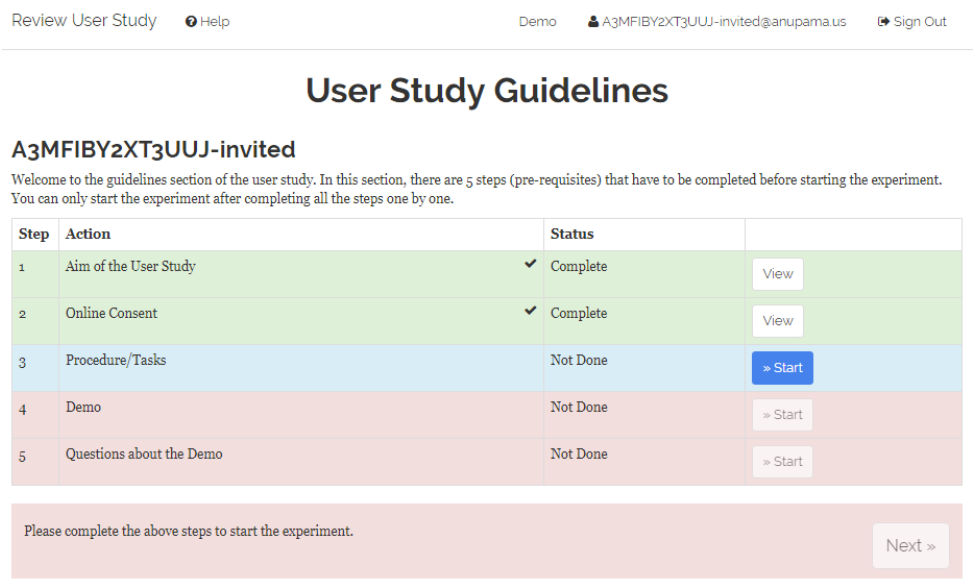


Figure 34. Completed (i) aim of the user study and (ii) online consent form

Next the participant is shown the procedure/tasks page (see Figure 35).

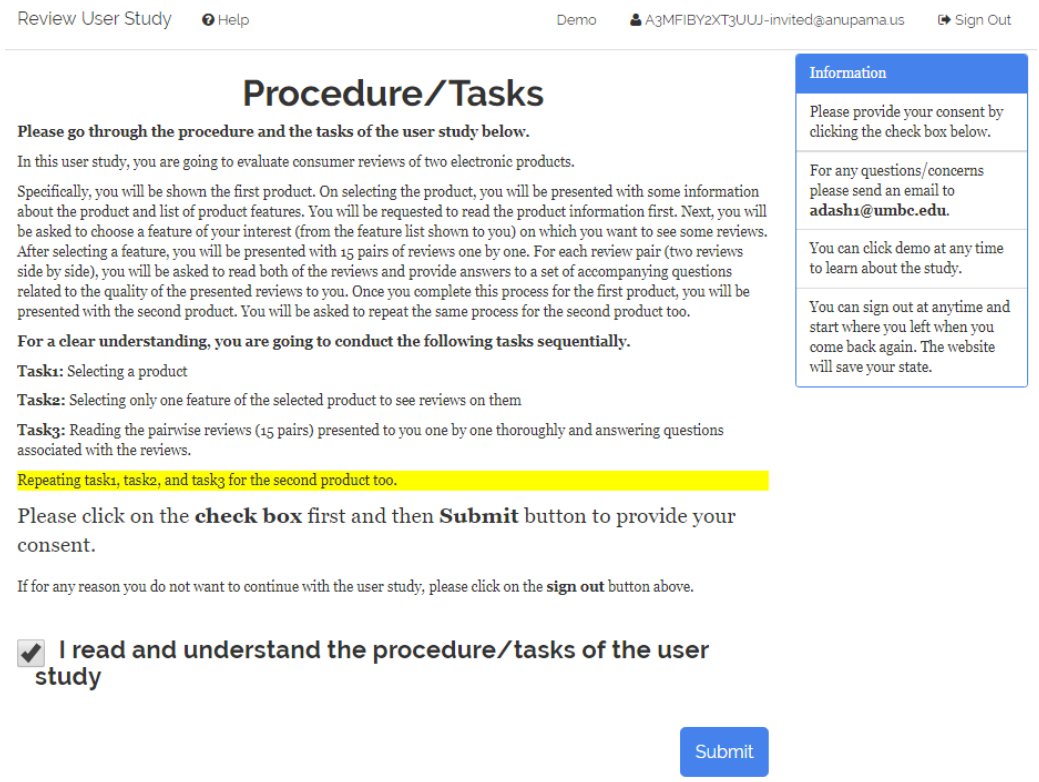


Figure 35. Procedure and tasks page

The participant is shown the demo video next (see **Figure 36**).

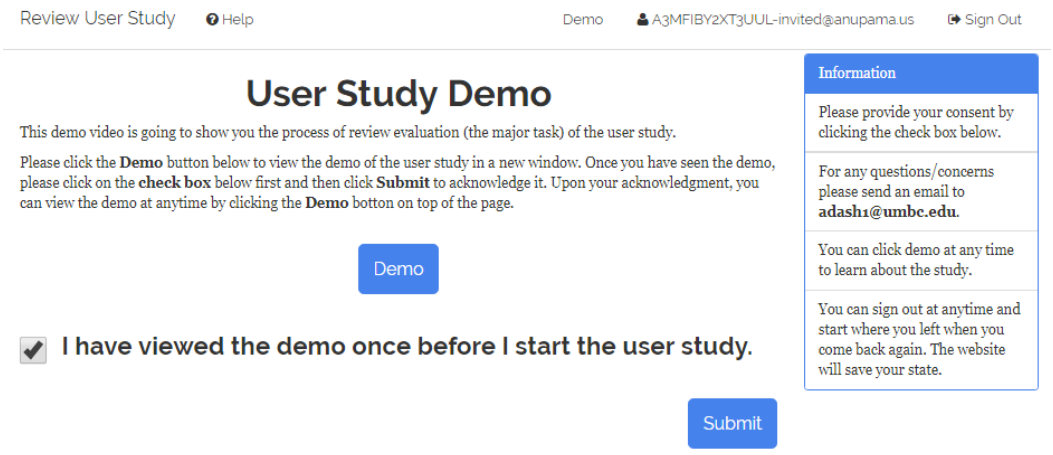


Figure 36. User study demo video page

A video that shows the user how the user study is performed is shown (see Figure 37).

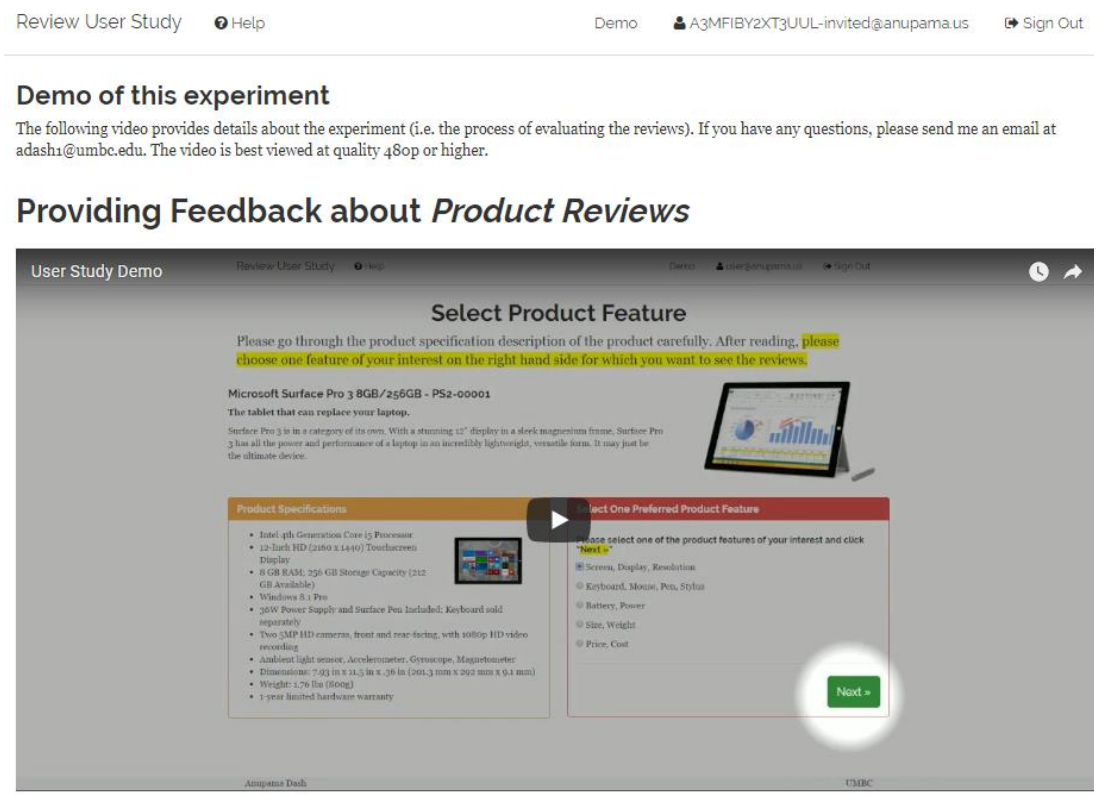


Figure 37. User study demo video player

Next the participant is asked two questions about the user study to make sure he/she understands the details about the study (see Figure 38).

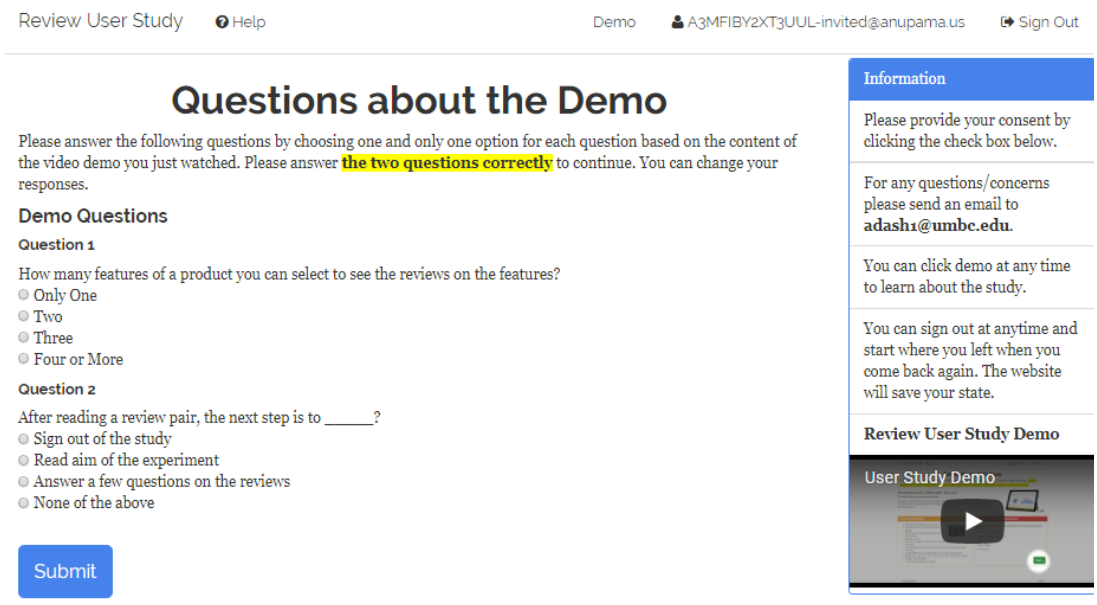


Figure 38. Questions about the demo

Once the participant answers the questions about the demo correctly he/she is shown the prerequisites completion screen and he/she is informed about the subsequent user study (Figure 39).

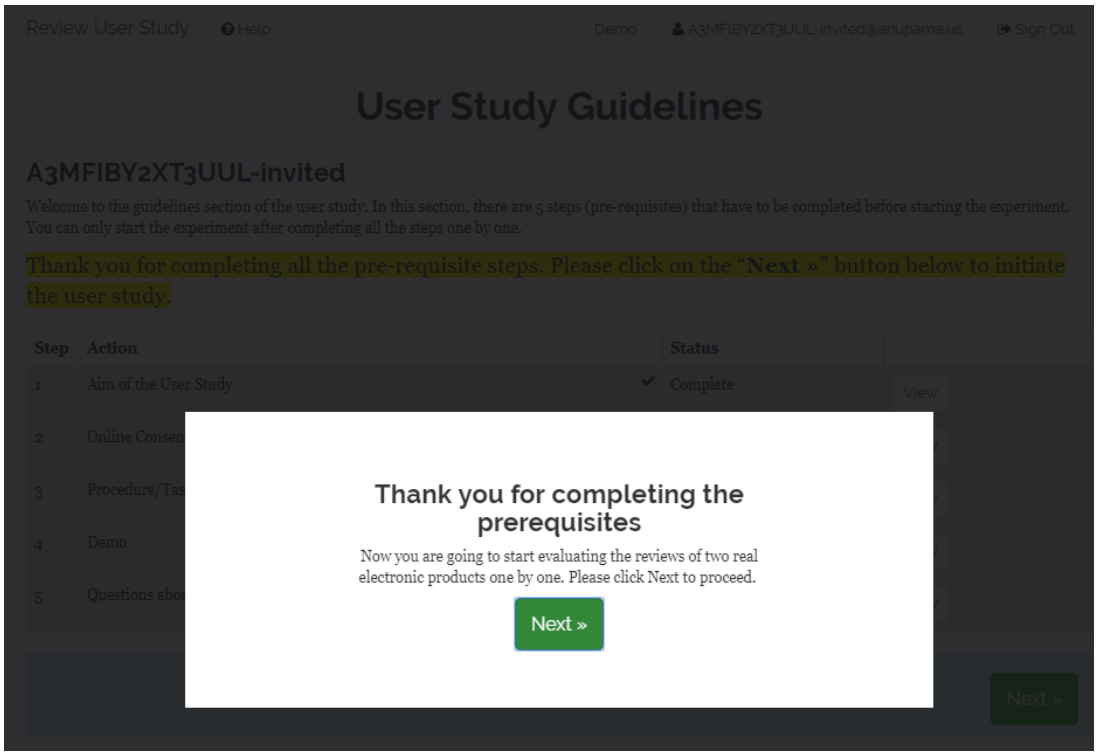


Figure 39. Participant completing the prerequisites and starting the study

The participant starts the user study by selecting the first product (see Figure 40).

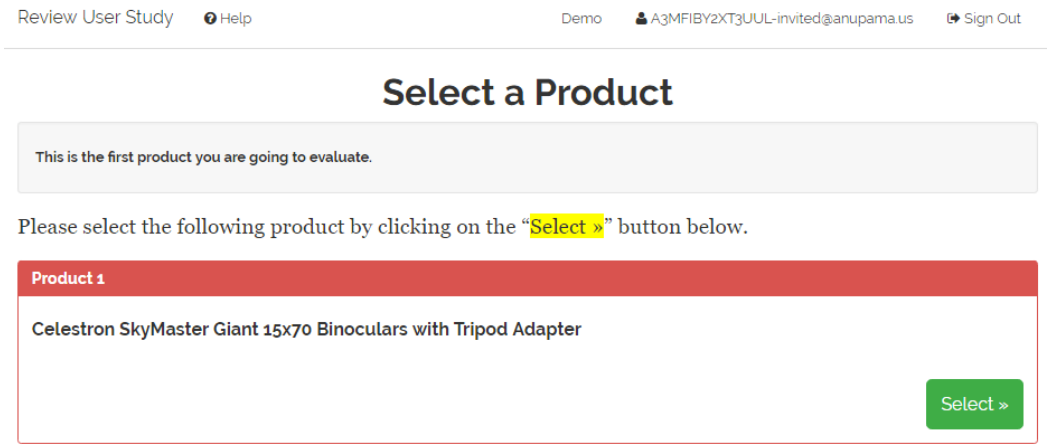


Figure 40. Participant selects the first product for the user study

Once the participant selects the product he/she is shown the details about the product and is requested to choose one feature that interest him the most (see Figure 41).

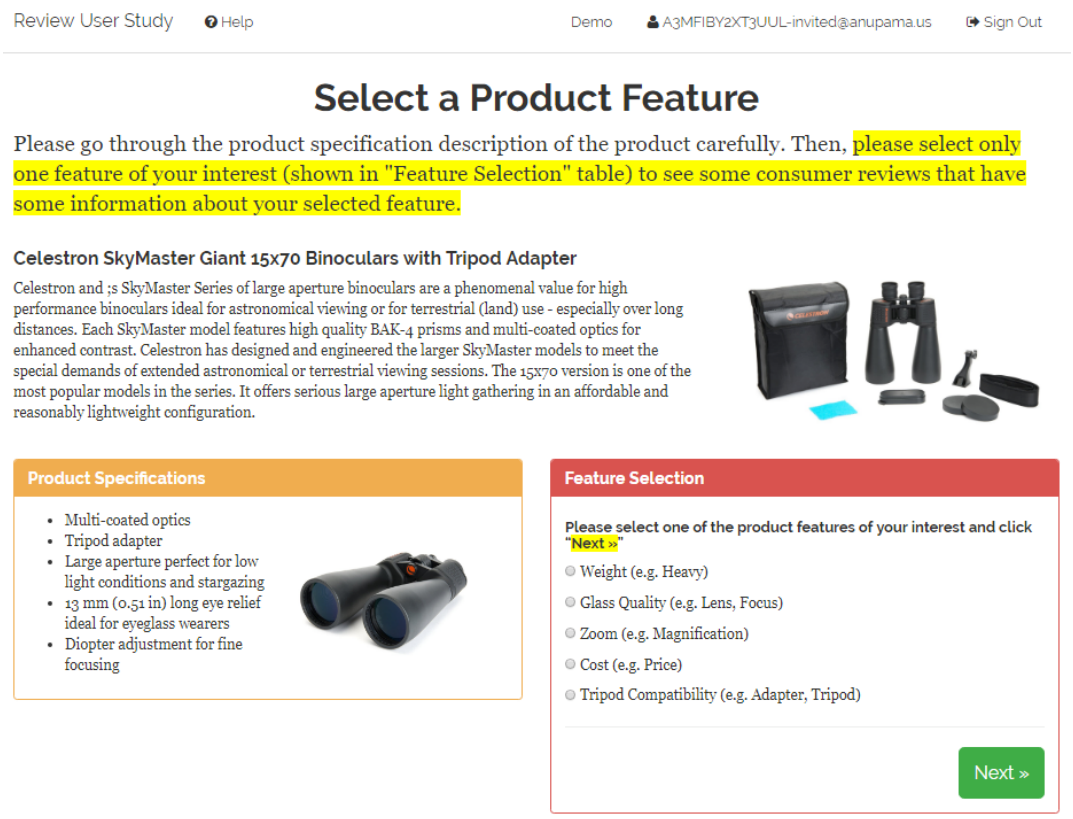


Figure 41. Product feature selection page in User Study

Once the participant selects a product feature he/she is show two reviews to compare (see Figure 42). The participant is asked a question about the two reviews to make sure he/she is paying attention to the reviews. The participant is also asked the three follow on questions about the knowledge, satisfaction, and relevancy. A progress bar is shown to the user on top to keep him aware of the progress in the user study. The participant is also shown the feature that he/she has selected in the product feature selection page.

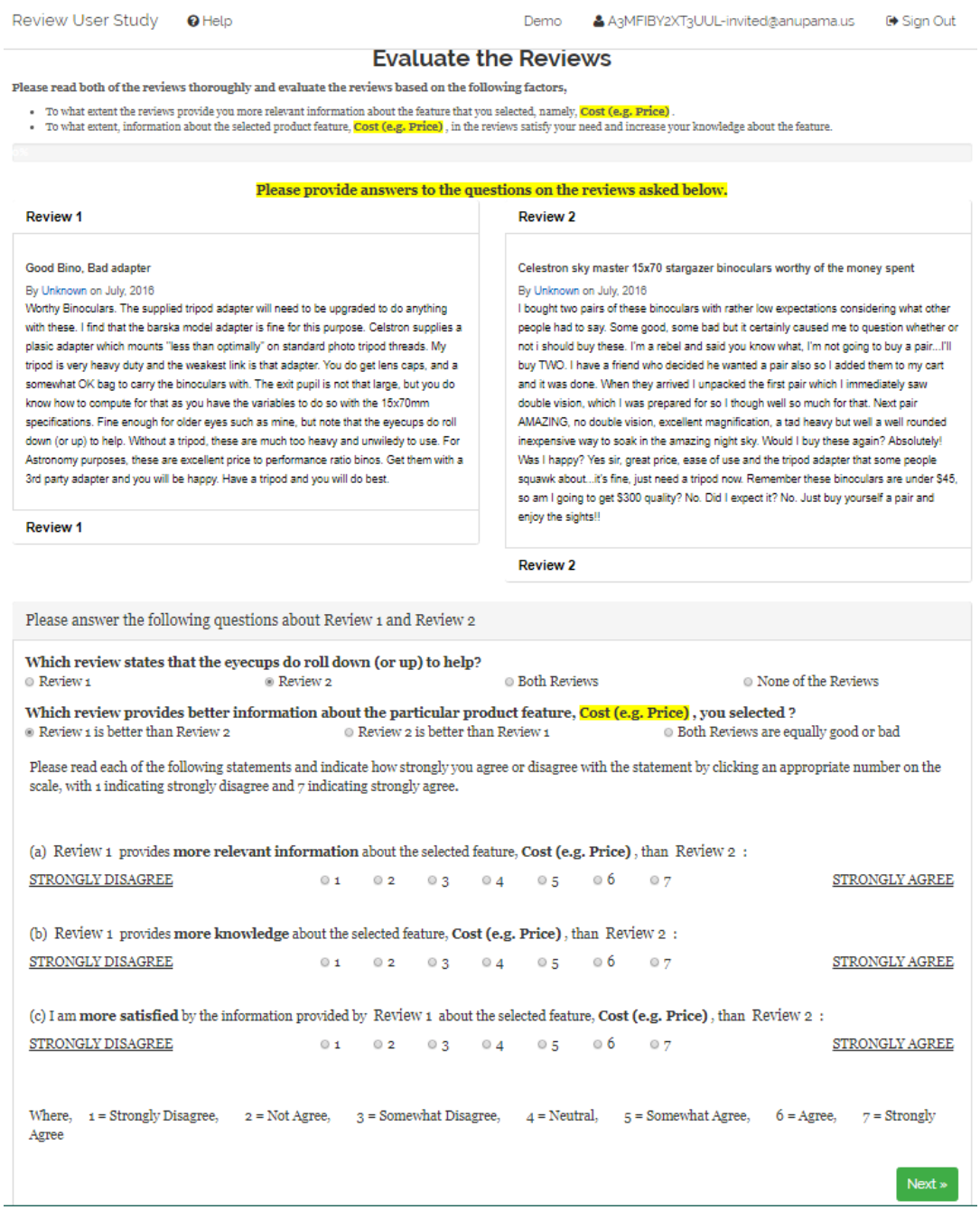


Figure 42. Product feature selection page in User Study

Once the participant completes all the pairwise comparison of the reviews he/she is shown the second and last product (see Figure 43) to provide feedback on.

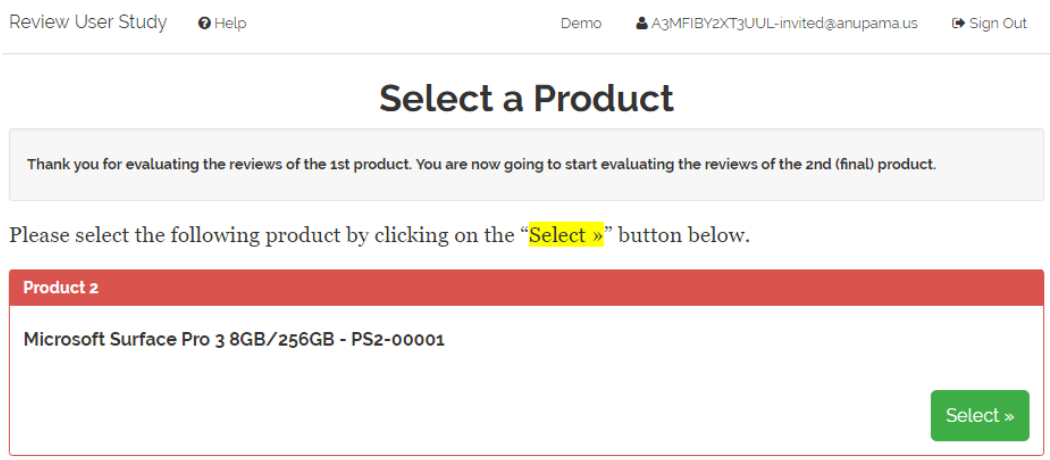


Figure 43. Product selection page in User Study

Similar to the first product the participant is again requested to choose a product feature from the second product that is of interest to him.

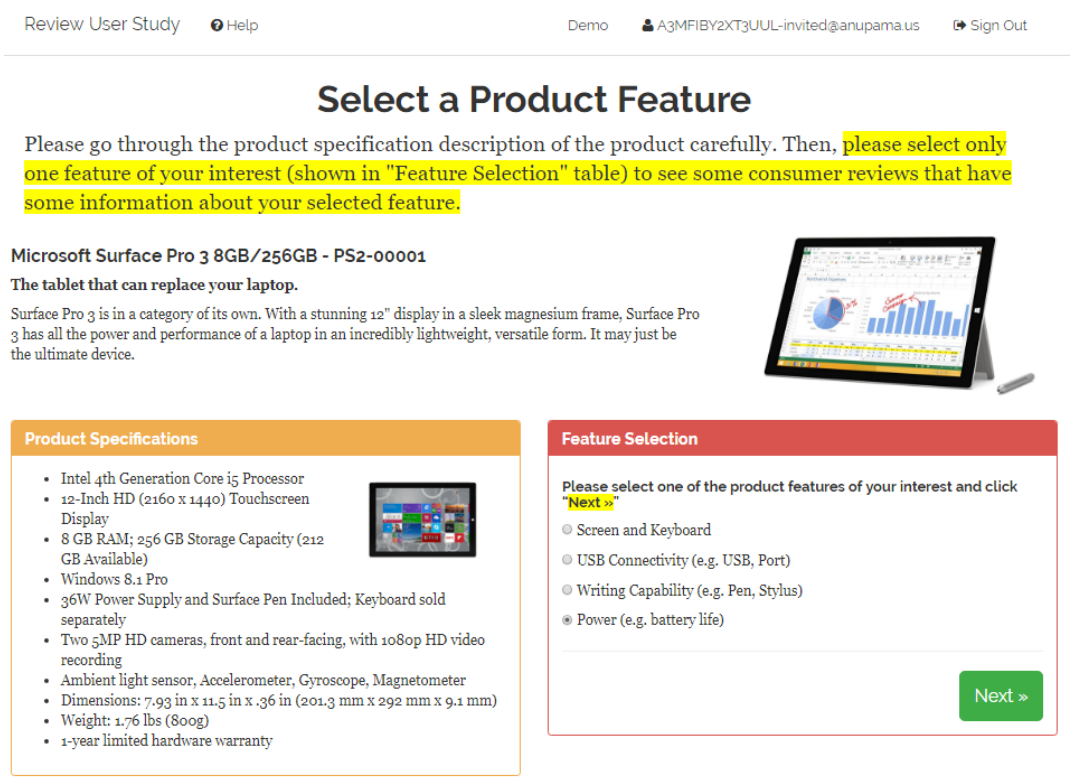


Figure 44. Product feature selection page in User Study

The participant is shown pair wise reviews from the second product and his/her feedback is collected. Once all the pairwise comparison is completed the participant

is shown the unique completion code that he/she enters in the Amazon mTurk page to provide evidence of user study completion (see Figure 45).

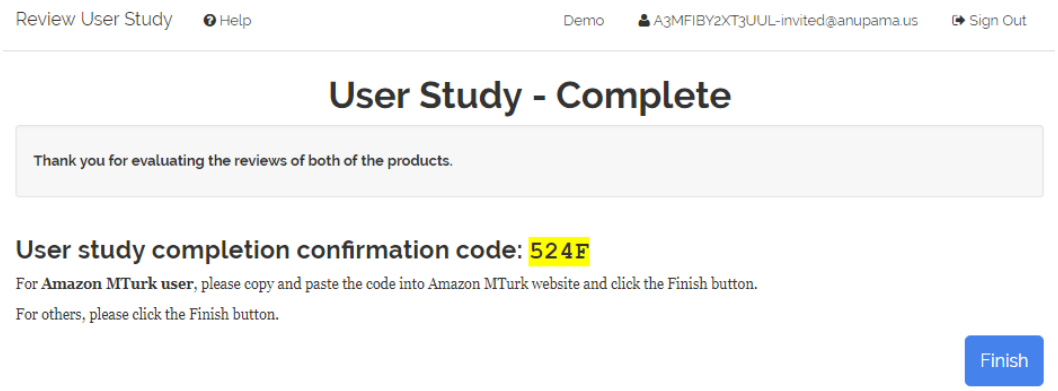


Figure 45. User Study completion page with unique confirmation code

The user is then requested to close the browser (see Figure 46).

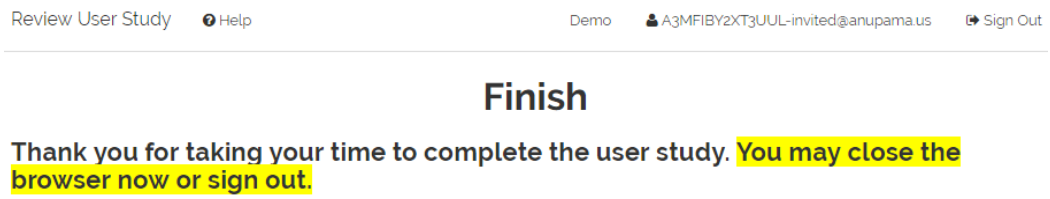


Figure 46. User Study finish page

Appendix C. Consistency and Distance Score for Product 1

Mturk Worker ID	LCR Class	Consistency	Helpfulness Ranking	LCR Ranking
A3MFIBY0000000	1	0.044	-0.333	0.733
A2EEL9Y0000000	2	0.102	-0.467	0.467
A2P4SHX0000000	1	0.104	-0.067	0.467
A1I9ZBV0000000	1	0.102	-0.067	0.467
A2EI0750000000	1	0.120	-0.067	0.733
A29VL3M0000000	2	0.122	-0.333	0.067
A36SM7Q0000000	2	0.083	-0.200	-0.067
A1945US0000000	2	0.142	-0.333	0.333
AX06ZUC0000000	2	0.120	-0.733	0.733
A35D31Q0000000	2	0.102	-0.600	0.333
A1NM7ZP0000000	2	0.120	-0.200	0.200
A2KLJKD0000000	2	0.102	-0.467	0.467
A11S8IA0000000	2	0.102	-0.467	0.467
A3DRSE70000000	2	0.121	-0.467	0.467
A1A3TGZ0000000	3	0.036	0.067	0.067
A2TT6FA0000000	1	0.060	-0.600	0.733
A2WGW5Y0000000	3	0.035	0.200	-0.067
A27W0250000000	3	0.049	-0.200	0.067
ADKKDK60000000	3	0.026	0.200	-0.333
A207IHY0000000	3	0.026	0.067	0.067
A1V6P1Z0000000	2	0.120	-0.067	-0.200
A2EEUQ00000000	2	0.120	-0.333	0.600
A3O2NOP0000000	1	0.165	0.200	-0.067
A5EU1AQ0000000	3	0.022	0.200	-0.067
A11NM7Z0000000	2	0.009	0.067	0.200
A1VNYP50000000	2	0.159	-0.200	-0.067
A2EOOF90000000	2	0.045	0.067	0.200
A19WXS10000000	1	0.063	-0.067	0.200
A3BI0AX0000000	2	0.091	0.467	-0.200
A2AAY4V0000000	1	0.031	-0.200	0.067
AHL33DD0000000	3	0.022	0.067	-0.200
A8F6JFG0000000	3	0.122	-0.600	0.733
A2OAZPR0000000	3	0.091	0.067	0.333
A1NKBXO0000000	1	0.044	-0.467	0.600
A2WNW8A0000000	2	0.040	0.067	-0.067
A28PS7T0000000	3	0.064	0.200	-0.333
A3EZ0H00000000	3	0.063	-0.467	0.333

Mturk Worker ID	LCR Class	Consistency	Helpfulness Ranking	LCR Ranking
A2SY4E70000000	3	0.035	0.200	-0.067
A3RU7AN0000000	1	0.049	-0.467	0.333
A3UUH360000000	3	0.159	-0.200	0.333
AJDXSXA0000000	2	0.018	-0.333	0.600
A1VW8Y70000000	2	0.083	0.333	-0.333
A1WRYUF0000000	2	0.102	-0.467	0.467
A31Z5TP0000000	3	0.035	0.067	0.067
A5LYLHG0000000	2	0.165	0.067	-0.067
A3C3RLZ0000000	2	0.156	-0.600	0.867
A1SBFOZ0000000	2	0.063	-0.200	0.467
A303MN10000000	2	0.190	-0.333	0.600
A1DR8T10000000	2	0.104	0.067	-0.067
A28A3HF0000000	2	0.083	-0.067	0.333
A9KSSJX0000000	2	0.044	-0.200	0.200
A27NN240000000	3	0.013	-0.067	0.200
A362CV80000000	2	0.063	-0.333	0.333
ASDKBXZ0000000	3	0.000	-0.067	0.200
A22051N0000000	2	0.075	-0.067	0.067
AKLV0WI0000000	2	0.059	0.067	-0.067
A1EI4NM0000000	2	0.039	-0.067	0.067
AROOCBM0000000	2	0.120	-0.467	0.467
ACCXC5V0000000	1	0.040	0.333	0.333
A10BH9P0000000	2	0.022	0.067	0.200
A2NT3OQ0000000	1	0.035	-0.067	0.733
A1XUZFD0000000	3	0.083	-1.000	0.600
A163J5T0000000	2	0.045	-0.200	0.467
A1LJKHC0000000	3	0.102	-0.467	0.867
ASTDBTV0000000	3	0.017	0.333	0.067

Appendix D. Consistency and Distance Score for Product 2

Mturk Worker ID	LCR Class	Consistency	Helpfulness Ranking	LCR Ranking
A3MFIBY0000000	2	0.044	-0.600	0.733
A2EEL9Y0000000	2	0.083	-0.733	0.867
A2P4SHX0000000	2	0.126	-0.467	0.600
A1I9ZBV0000000	2	0.102	0.067	0.333
A2EI0750000000	2	0.083	-0.200	0.067
A29VL3M0000000	2	0.121	-0.600	0.200
A36SM7Q0000000	2	0.063	-0.733	0.867
A1945US0000000	2	0.063	-0.333	0.467
AX06ZUC0000000	3	0.120	-0.333	0.733
A35D31Q0000000	3	0.122	-0.867	0.467
A1NM7ZP0000000	3	0.139	-0.600	0.733
A2KLJKD0000000	3	0.102	-0.333	0.733
A11S8IA0000000	1	0.120	-0.467	0.467
A3DRSE70000000	1	0.122	-0.333	0.333
A1A3TGZ0000000	1	0.000	-0.600	0.600
A2TT6FA0000000	2	0.022	-0.067	-0.067
A2WGW5Y0000000	2	0.000	-0.200	0.067
A27W0250000000	2	0.009	-0.467	0.067
ADKKDK60000000	2	0.000	-0.200	0.067
A207IHY0000000	3	0.026	-0.333	0.200
A1V6P1Z0000000	2	0.059	-0.600	0.467
A2EEUQ00000000	2	0.102	-0.467	0.600
A3O2NOP0000000	2	0.159	-0.067	0.200
A5EU1AQ0000000	2	0.044	-0.733	0.600
A11NM7Z0000000	2	0.000	-0.200	0.067
A1VNYP50000000	3	0.119	0.067	0.067
A2EOOF90000000	2	0.044	-0.467	0.600
A19WXS10000000	3	0.075	-0.333	0.200
A3BI0AX0000000	1	0.105	-0.600	0.600
A2AAY4V0000000	2	0.000	-0.200	0.067
AHL33DD0000000	2	0.091	-0.200	-0.200
A8F6JFG0000000	1	0.139	-0.467	0.200
A2OAZPR0000000	1	0.111	-0.600	0.600
A1NKBXO0000000	3	0.031	0.067	0.333
A2WNW8A0000000	2	0.013	-0.867	0.733
A28PS7T0000000	3	0.072	0.733	-0.333
A3EZ0H00000000	2	0.036	0.333	0.067

Mturk Worker ID	LCR Class	Consistency	Helpfulness Ranking	LCR Ranking
A2SY4E70000000	2	0.000	-0.200	0.067
A3RU7AN0000000	2	0.000	-0.200	0.067
A3UUH360000000	1	0.104	-0.067	0.333
AJDXSXA0000000	3	0.035	-0.333	0.467
A1VW8Y70000000	3	0.122	-0.200	0.333
A1WRYUF0000000	2	0.009	-0.200	0.067
A31Z5TP0000000	2	0.022	-1.000	0.600
A5LYLHG0000000	2	0.118	-0.333	-0.067
A3C3RLZ0000000	2	0.100	0.067	-0.467
A1SBFOZ0000000	2	0.121	-0.467	0.067
A303MN10000000	2	0.063	-0.067	0.200
A1DR8T10000000	2	0.083	-0.200	-0.200
A28A3HF0000000	2	0.044	0.467	-0.600
A9KSSJX0000000	2	0.022	0.200	-0.067
A27NN240000000	3	0.081	0.067	-0.200
A362CV80000000	2	0.102	-0.333	0.467
ASDKBXZ0000000	2	0.022	-0.733	0.333
A22051N0000000	2	0.009	-0.467	0.333
AKLV0WI0000000	2	0.040	-0.200	0.333
A1EI4NM0000000	2	0.000	-0.200	0.067
AROOCBM0000000	3	0.053	-0.867	0.467
ACCXC5V0000000	2	0.000	-0.200	0.067
A10BH9P0000000	2	0.000	-0.200	0.067
A2NT3OQ0000000	3	0.083	0.200	-0.333
A1XUZFD0000000	1	0.063	-0.467	0.467
A163J5T0000000	2	0.000	-0.200	0.067
A1LJKHC0000000	2	0.081	0.200	-0.600
ASTDBTV0000000	3	0.022	-0.467	0.333

Appendix E. T-Test results

This appendix presents the t-test results of the user study.

Table 29. T-Test for primary level user ranking of product 1		
	Variable 1	Variable 2
Mean	0.572881	0.380791
Variance	0.0227	0.023916
Observations	59	59
Pearson Correlation	-0.71667	
Hypothesized Mean Difference	0	
Df	58	
t Stat	5.216166	
P(T<=t) one-tail	1.28E-06	
t Critical one-tail	1.671553	

Table 30. T-Test for primary level user ranking of product 2		
	Variable 1	Variable 2
Mean	0.650256	0.377436
Variance	0.027088	0.029205
Observations	62	62
Pearson Correlation	-0.75654	
Hypothesized Mean Difference	0	
Df	61	
t Stat	6.995913	
P(T<=t) one-tail	9.46E-10	
t Critical one-tail	1.669013	

Table 31. T-Test for user relevance ranking of product 1		
	Variable 1	Variable 2
Mean	0.50991	0.423423
Variance	0.01635	0.016844
Observations	37	37
Pearson Correlation	-0.68765	
Hypothesized Mean Difference	0	
Df	36	
t Stat	2.222762	
P(T<=t) one-tail	0.0163	
t Critical one-tail	1.688298	

Table 32. T-Test for user relevance ranking of product 2		
	Variable 1	Variable 2
Mean	0.62439	0.417886
Variance	0.020835	0.024894
Observations	41	41
Pearson Correlation	-0.79537	
Hypothesized Mean Difference	0	
Df	40	
t Stat	4.618793	
P(T<=t) one-tail	1.98E-05	
t Critical one-tail	1.683851	

Table 33. T-Test for user knowledge ranking of product 1		
	Variable 1	Variable 2
Mean	0.557333	0.424
Variance	0.031067	0.037363
Observations	25	25
Pearson Correlation	-0.8275	
Hypothesized Mean Difference	0	
Df	24	
t Stat	1.887016	
P(T<=t) one-tail	0.035656	
t Critical one-tail	1.710882	

Table 34. T-Test for user knowledge ranking of product 2		
	Variable 1	Variable 2
Mean	0.633333	0.351111
Variance	0.035172	0.037987
Observations	30	30
Pearson Correlation	-0.81759	
Hypothesized Mean Difference	0	
Df	29	
t Stat	4.239755	
P(T<=t) one-tail	0.000104	
t Critical one-tail	1.699127	

Table 35. T-Test for user satisfaction ranking of product 1		
	Variable 1	Variable 2
Mean	0.5125	0.452083
Variance	0.010735	0.01168
Observations	32	32
Pearson Correlation	-0.57857	
Hypothesized Mean Difference	0	
Df	31	
t Stat	1.8172	
P(T<=t) one-tail	0.039433	
t Critical one-tail	1.695519	

Table 36. T-Test for user satisfaction ranking of product 2		
	Variable 1	Variable 2
Mean	0.598291	0.452991
Variance	0.023974	0.023668
Observations	39	39
Pearson Correlation	-0.84063	
Hypothesized Mean Difference	0	
Df	38	
t Stat	3.064235	
P(T<=t) one-tail	0.002001	
t Critical one-tail	1.685954	

Appendix F. Consistency and Distance Score for Product 1:

Relevance

UserID	LCR Class	Consistency	Helpfulness Ranking	LCR Ranking
1	1	0.125309	0.666667	0.133333
2	2	0.100748	0.6	0.266667
3	3	0.099559	0.466667	0.466667
4	3	0.038188	0.466667	0.466667
5	3	0.098901	0.666667	0.4
6	3	0.026174	0.4	0.666667
7	3	0.026754	0.4	0.533333
8	3	0.043482	0.4	0.533333
9	2	0.009534	0.533333	0.333333
10	2	0.145382	0.6	0.4
11	1	0.061479	0.466667	0.466667
12	2	0.137395	0.266667	0.6
13	1	0.080194	0.666667	0.4
14	3	0.030185	0.466667	0.6
15	1	0.055837	0.666667	0.266667
16	2	0.056966	0.466667	0.533333
17	3	0.141443	0.333333	0.733333
18	3	0.09734	0.866667	0.2
19	3	0.051516	0.4	0.533333
20	1	0.02183	0.666667	0.4
21	2	0.014497	0.533333	0.333333
22	2	0.101258	0.6	0.4
23	3	0.054007	0.466667	0.466667
24	2	0.096293	0.6	0.266667
25	2	0.085504	0.466667	0.4
26	2	0.08948	0.6	0.4
27	3	0.02587	0.466667	0.466667
28	2	0.030898	0.466667	0.4
29	3	0	0.533333	0.4
30	2	0.138044	0.6	0.4
31	2	0.077036	0.466667	0.533333
32	2	0.037754	0.466667	0.533333
33	1	0.142877	0.266667	0.4
34	2	0.063051	0.466667	0.4
35	1	0.04345	0.4	0.266667
36	2	0.092089	0.666667	0.2
37	3	0.082892	0.333333	0.466667

Appendix G. Consistency and Distance Score for Product 1: Knowledge

UserID	LCR Class	Consistency	Helpfulness Ranking	LCR Ranking
1	3	0	0.466667	0.466667
2	3	0	0.533333	0.4
3	3	0	0.866667	0.066667
4	3	0	0.466667	0.6
5	3	0	0.466667	0.6
6	3	0	0.4	0.666667
7	2	0	0.533333	0.333333
8	1	0.043414	0.8	0.4
9	3	0	0.533333	0.533333
10	1	0	0.4	0.533333
11	2	0	0.733333	0.266667
12	3	0	0.866667	0.066667
13	3	0.129122	0.4	0.533333
14	1	0	0.8	0.4
15	2	0	0.266667	0.733333
16	3	0.143448	0.466667	0.466667
17	3	0	0.466667	0.466667
18	3	0	0.466667	0.6
19	2	0	0.733333	0.2
20	2	0	0.6	0.4
21	1	0	0.733333	0.2
22	2	0	0.466667	0.4
23	1	0.129646	0.466667	0.2
24	2	0	0.733333	0.266667
25	3	0	0.266667	0.8

Appendix H. Consistency and Distance Score for Product 1: Satisfaction

UserID	LCR Class	Consistency	Helpfulness Ranking	LCR Ranking
1	2	0.081199	0.666667	0.466667
2	3	0.093834	0.466667	0.466667
3	3	0.034445	0.533333	0.4
4	3	0.104665	0.533333	0.533333
5	3	0.031735	0.466667	0.6
6	3	0.026754	0.4	0.533333
7	3	0.037454	0.4	0.533333
8	2	0.018674	0.6	0.266667
9	1	0.069628	0.6	0.466667
10	2	0.143375	0.533333	0.466667
11	1	0.075351	0.533333	0.533333
12	3	0.051062	0.466667	0.6
13	3	0.091515	0.533333	0.4
14	2	0.098992	0.466667	0.533333
15	3	0.05058	0.4	0.666667
16	3	0.066223	0.8	0.266667
17	3	0.051516	0.4	0.533333
18	1	0.055979	0.666667	0.4
19	2	0.01592	0.6	0.266667
20	3	0.061311	0.466667	0.466667
21	2	0.139797	0.533333	0.333333
22	2	0.095911	0.6	0.4
23	3	0.042724	0.4	0.533333
24	2	0.065008	0.533333	0.466667
25	3	0	0.533333	0.4
26	2	0.132366	0.6	0.4
27	2	0.102718	0.466667	0.533333
28	2	0.053228	0.466667	0.533333
29	2	0.092397	0.466667	0.4
30	1	0.063977	0.4	0.266667
31	2	0.1261	0.6	0.266667
32	3	0.098593	0.266667	0.533333

Appendix I. Consistency and Distance Score for Product 2:

Relevance

UserID	LCR Class	Consistency	Helpfulness Ranking	LCR Ranking
1	2	0.065274	0.8	0.266667
2	2	0.123008	0.8	0.266667
3	2	0.120416	0.533333	0.266667
4	2	0.107245	0.866667	0.066667
5	3	0.116869	0.533333	0.266667
6	1	0	0.8	0.2
7	2	0.037379	0.533333	0.533333
8	2	0	0.6	0.466667
9	2	0.0088	0.733333	0.466667
10	2	0	0.6	0.466667
11	3	0.040623	0.666667	0.4
12	2	0.039959	0.8	0.266667
13	2	0	0.6	0.466667
14	3	0.101532	0.6	0.466667
15	1	0.12546	0.866667	0.133333
16	2	0	0.6	0.466667
17	2	0.031516	0.733333	0.466667
18	3	0.076649	0.466667	0.333333
19	2	0	0.6	0.466667
20	2	0.026581	0.333333	0.6
21	2	0	0.6	0.466667
22	2	0	0.6	0.466667
23	3	0.026958	0.6	0.333333
24	2	0.022807	0.6	0.466667
25	2	0.054664	0.866667	0.2
26	2	0.100975	0.466667	0.466667
27	2	0.124433	0.6	0.6
28	2	0.057618	0.266667	0.8
29	2	0.029937	0.466667	0.466667
30	2	0.039959	0.666667	0.4
31	2	0.014497	0.666667	0.533333
32	2	0	0.6	0.466667
33	2	0.048542	0.733333	0.333333
34	2	0	0.6	0.466667
35	2	0	0.6	0.466667
36	2	0	0.6	0.466667
37	3	0.141235	0.4	0.666667
38	1	0.125434	0.733333	0.266667

39	2	0	0.6	0.466667
40	2	0.044323	0.4	0.8
41	3	0.080857	0.866667	0.2

Appendix J. Consistency and Distance Score for

Product 2: Knowledge

UserID	LCR Class	Consistency	Helpfulness Ranking	LCR Ranking
1	1	0	0.866667	0.266667
2	2	0	0.666667	0.266667
3	2	0	0.733333	0.2
4	2	0	0.933333	0.266667
5	2	0	0.733333	0.2
6	3	0	0.666667	0.4
7	2	0	0.6	0.333333
8	2	0	0.733333	0.2
9	2	0	0.733333	0.2
10	2	0.090507	0.6	0.6
11	3	0.083316	0.466667	0.466667
12	2	0	0.4	0.4
13	3	0	0.2	0.6
14	2	0	0.2	0.733333
15	2	0	0.733333	0.2
16	2	0	0.733333	0.2
17	3	0	0.666667	0.266667
18	2	0	0.733333	0.2
19	2	0	0.666667	0.4
20	2	0	0.466667	0.333333
21	2	0	0.8	0.266667
22	2	0	0.333333	0.866667
23	2	0	0.533333	0.4
24	2	0	0.733333	0.2
25	3	0.131042	0.933333	0.133333
26	2	0	0.733333	0.2
27	2	0	0.733333	0.2
28	2	0	0.733333	0.2
29	2	0	0.466667	0.733333
30	3	0	0.466667	0.6

Appendix K. Consistency and Distance Score for

Product 2: Satisfaction

UserID	LCR Class	Consistency	Helpfulness Ranking	LCR Ranking
1	2	0.076689	0.6	0.6
2	2	0.118565	0.533333	0.4
3	1	0.022807	0.866667	0.266667
4	2	0.037379	0.533333	0.533333
5	2	0	0.6	0.466667
6	2	0.0088	0.733333	0.466667
7	2	0	0.6	0.466667
8	3	0.040623	0.666667	0.4
9	2	0.148732	0.8	0.266667
10	2	0.0725	0.666667	0.133333
11	2	0	0.6	0.466667
12	3	0.128363	0.6	0.466667
13	3	0.071265	0.466667	0.6
14	1	0.148773	0.8	0.2
15	2	0	0.6	0.466667
16	2	0.0088	0.533333	0.533333
17	3	0.06578	0.6	0.333333
18	2	0	0.6	0.466667
19	3	0.121683	0.2	0.6
20	2	0.041012	0.333333	0.733333
21	2	0	0.6	0.466667
22	2	0	0.6	0.466667
23	3	0.033385	0.6	0.333333
24	2	0.037379	0.6	0.466667
25	2	0.054664	0.866667	0.2
26	2	0.057618	0.266667	0.8
27	2	0.022807	0.266667	0.666667
28	2	0.036424	0.8	0.266667
29	2	0.037379	0.733333	0.466667
30	2	0	0.6	0.466667
31	2	0.032321	0.733333	0.333333
32	2	0	0.6	0.466667
33	2	0	0.6	0.466667
34	2	0	0.6	0.466667
35	3	0.125703	0.466667	0.6
36	1	0.119979	0.666667	0.333333
37	2	0	0.6	0.466667
38	2	0.044323	0.4	0.8

39	3	0.102422	0.8	0.266667

References

[1] F. R. Jiménez and N. A. Mendoza, “Too Popular to Ignore: The Influence of Online Reviews on Purchase Intentions of Search and Experience Products,” *J. Interact. Mark.*, vol. 27, no. 3, pp. 226–235, Aug. 2013.

[2] T. Hennig-Thurau, K. P. Gwinner, G. Walsh, and D. D. Gremler, “Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet?,” *J. Interact. Mark.*, vol. 18, no. 1, pp. 38–52, Dec. 2004.

[3] “Under the Influence: Consumer Trust In Advertising.” [Online]. Available: http://www.nielsen.com/us/en/insights/news/2013/under-the-influence-consumer-trust-in-advertising.html. [Accessed: 25-Oct-2017].

[4] P.-Y. Chen, S. Dhanasobhon, and M. D. Smith, “All Reviews Are Not Created Equal: The Disaggregate Impact of Reviews and Reviewers at Amazon.Com,” *SSRN ELibrary*, May 2008.

[5] P. Dabholkar, “Factors Influencing Consumer Choice of a ‘Rating Web Site’: An Experimental Investigation of an Online Interactive Decision Aid,” *J. Mark. Theory Pract.*, vol. 14, no. 4, pp. 259–273, Oct. 2006.

[6] C. Forman, A. Ghose, and B. Wiesenfeld, “Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets,” *Inf. Syst. Res.*, vol. 19, no. 3, pp. 291–313, 2008.

[7] W. Duan, B. Gu, and A. B. Whinston, “Do online reviews matter? — An empirical investigation of panel data,” *Decis. Support Syst.*, vol. 45, no. 4, pp. 1007–1016, Nov. 2008.

[8] N. Kumar and I. Benbasat, “Research Note: The Influence of Recommendations and Consumer Reviews on Evaluations of Websites,” *Info Sys Res.*, vol. 17, no. 4, pp. 425–439, Dec. 2006.

[9] J. A. Chevalier and D. Mayzlin, “The Effect of Word of Mouth on Sales: Online Book Reviews,” *J. Mark. Res.*, vol. 43, no. 3, pp. 345–354, Aug. 2006.

[10] H.-J. Min and J. C. Park, “Identifying Helpful Reviews Based on Customer’s Mentions About Experiences,” *Expert Syst Appl*, vol. 39, no. 15, pp. 11830–11838, Nov. 2012.

[11] Q. Cao, W. Duan, and Q. Gan, “Exploring Determinants of Voting for the ‘Helpfulness’ of Online User Reviews: A Text Mining Approach,” *Decis Support Syst*, vol. 50, no. 2, pp. 511–521, Jan. 2011.

[12] D.-H. Park and J. Lee, “eWOM overload and its effect on consumer behavioral intention depending on consumer involvement,” *Electron. Commer. Res. Appl.*, vol. 7, no. 4, pp. 386–398, Winter 2008.

[13] J. Tang, H. Gao, X. Hu, and H. Liu, “Context-aware Review Helpfulness Rating Prediction,” in *Proceedings of the 7th ACM Conference on Recommender Systems*, New York, NY, USA, 2013, pp. 1–8.

[14] Y. Lu, P. Tsaparas, A. Ntoulas, and L. Polanyi, “Exploiting Social Context for Review Quality Prediction,” in *Proceedings of the 19th International Conference on World Wide Web*, New York, NY, USA, 2010, pp. 691–700.

[15] A. Ghose and P. G. Ipeirotis, “Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics,” *IEEE Trans. Knowl. Data Eng.*, vol. 99, no. 1, 2009.
[16] D. Weathers, S. D. Swain, and V. Grover, “Can online product reviews be more helpful? Examining characteristics of information content by product type,” *Decis. Support Syst.*, vol. 79, pp. 12–23, Nov. 2015.
[17] A. Y. K. Chua and S. Banerjee, “Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality,” *Comput. Hum. Behav.*, vol. 54, pp. 547–554, Jan. 2016.
[18] S. Moghaddam, M. Jamali, and M. Ester, “ETF: Extended Tensor Factorization Model for Personalizing Prediction of Review Helpfulness,” in *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, New York, NY, USA, 2012, pp. 163–172.
[19] L. Xia and N. N. Bechwati, “Word of Mouse: The Role of Cognitive Personalization in Online Consumer Reviews,” *J. Interact. Advert.*, vol. 9, no. 1, 2008.
[20] H.-Y. Ha, “The Effects of Consumer Risk Perception on Pre-purchase Information in Online Auctions: Brand, Word-of-Mouth, and Customized Information,” *J. Comput.-Mediat. Commun.*, vol. 8, no. 1, pp. 0–0, 2002.
[21] C. C. Chen and Y.-D. Tseng, “Quality Evaluation of Product Reviews Using an Information Quality Framework,” *Decis Support Syst*, vol. 50, no. 4, pp. 755–768, Mar. 2011.
[22] S.-M. Kim, P. Pantel, T. Chklovski, and M. Pennacchiotti, “Automatically Assessing Review Helpfulness,” in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, 2006, pp. 423–430.
[23] G. Adomavicius and A. Tuzhilin, “Context-Aware Recommender Systems,” in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer US, 2011, pp. 217–253.
[24] S. M. Mudambi and D. Schuff, “What Makes a Helpful Review? A Study of Customer Reviews on Amazon.com,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2175066, 2010.
[25] K. Berezina, A. Bilgihan, C. Cobanoglu, and F. Okumus, “Understanding Satisfied and Dissatisfied Hotel Customers: Text Mining of Online Hotel Reviews,” *J. Hosp. Mark. Manag.*, vol. 25, no. 1, pp. 1–24, Jan. 2016.
[26] “Exploring the Role of Preference Heterogeneity and Causal Attribution in Online Ratings Dynamics - ASIA MARKETING JOURNAL - Korean Marketing Association : Journal article - DBpia.” [Online]. Available: http://www.dbpia.co.kr/Journal/ArticleDetail/NODE02465987. [Accessed: 07-Sep-2017].
[27] N. Hu, L. Liu, and J. J. Zhang, “Do online reviews affect product sales? The role of reviewer characteristics and temporal effects,” *Inf. Technol. Manag.*, vol. 9, no. 3, pp. 201–214, Mar. 2008.
[28] N. Hu, N. S. Koh, and S. K. Reddy, “Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales,” *Decis. Support Syst.*, vol. 57, pp. 42–53, Jan. 2014.

[29] Y.-H. Chen and S. Barnes, “Initial trust and online buyer behaviour,” *Ind. Manag. Data Syst.*, vol. 107, no. 1, pp. 21–36, Jan. 2007.
[30] J.-J. Wu and Y.-S. Chang, “Towards understanding members’ interactivity, trust, and flow in online travel community,” *Ind. Manag. Data Syst.*, vol. 105, no. 7, pp. 937–954, Sep. 2005.
[31] W. Duan, B. Gu, and A. B. Whinston, “The dynamics of online word-of-mouth and product sales--An empirical investigation of the movie industry,” *J. Retail.*, vol. 84, no. 2, pp. 233–242, Jun. 2008.
[32] E.-J. Lee and S. Y. Shin, “When do consumers buy online product reviews? Effects of review quality, product type, and reviewer’s photo,” *Comput. Hum. Behav.*, vol. 31, pp. 356–366, Feb. 2014.
[33] D.-H. Park, J. Lee, and I. Han, “The Effect of On-Line Consumer Reviews on Consumer Purchasing Intention: The Moderating Role of Involvement,” *Int. J. Electron. Commer.*, vol. 11, pp. 125–148, Jul. 2007.
[34] J. A. Chevalier and D. Mayzlin, “The Effect of Word of Mouth on Sales: Online Book Reviews,” National Bureau of Economic Research, Working Paper 10148, Dec. 2003.
[35] Y. Liu, “Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,” *J. Mark.*, vol. 70, no. 3, pp. 74–89, Jul. 2006.
[36] A. H. Huang, K. Chen, D. C. Yen, and T. P. Tran, “A study of factors that contribute to online review helpfulness,” *Comput. Hum. Behav.*, vol. 48, pp. 17–27, Jul. 2015.
[37] M. Salehan and D. J. Kim, “Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics,” *Decis. Support Syst.*, vol. 81, pp. 30–40, Jan. 2016.
[38] F. Zhu and X. (Michael) Zhang, “Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics,” *J. Mark.*, vol. 74, no. 2, pp. 133–148, Mar. 2010.
[39] Y. Pan and J. Q. Zhang, “Born Unequal: A Study of the Helpfulness of User-Generated Product Reviews,” *J. Retail.*, vol. 87, no. 4, pp. 598–612, Dec. 2011.
[40] P. Racherla and W. Friske, “Perceived ‘usefulness’ of online consumer reviews: An exploratory investigation across three services categories,” *Electron. Commer. Res. Appl.*, vol. 11, no. 6, pp. 548–559, Nov. 2012.
[41] L. Kwok and K. L. Xie, “Factors contributing to the helpfulness of online hotel reviews: Does manager response play a role?,” *Int. J. Contemp. Hosp. Manag.*, vol. 28, no. 10, pp. 2156–2177, Oct. 2016.
[42] H. Baek, J. Ahn, and Y. Choi, “Helpfulness of Online Consumer Reviews: Readers’ Objectives and Review Cues,” *Int. J. Electron. Commer.*, vol. 17, no. 2, pp. 99–126, Dec. 2012.
[43] P. F. Wu, “In Search of Negativity Bias: An Empirical Study of Perceived Helpfulness of Online Reviews,” *Psychol. Mark.*, vol. 30, no. 11, pp. 971–984, Nov. 2013.
[44] J. Q. Zhang, G. Craciun, and D. Shin, “When does electronic word-of-mouth matter? A study of consumer product reviews,” *J. Bus. Res.*, vol. 63, no. 12, pp. 1336–1341, Dec. 2010.

[45] M. Li, L. Huang, C.-H. Tan, and K.-K. Wei, “Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features,” *Int. J. Electron. Commer.*, vol. 17, no. 4, pp. 101–136, Jul. 2013.
[46] D. Yin, S. Bond, and H. Zhang, “Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2263561, May 2013.
[47] A. Felbermayr and A. Nanopoulos, “The Role of Emotions for the Perceived Usefulness in Online Customer Reviews,” *J. Interact. Mark.*, vol. 36, pp. 60–76, Nov. 2016.
[48] “How Do Expressed Emotions Affect the Helpfulness of a Product Review? Evidence from Reviews Using Latent Semantic Analysis: International Journal of Electronic Commerce: Vol 20, No 1.” [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/10864415.2016.1061471. [Accessed: 24-Jan-2018].
[49] S. G. Moore, “Attitude Predictability and Helpfulness in Online Reviews: The Role of Explained Actions and Reactions,” *J. Consum. Res.*, vol. 42, no. 1, pp. 30–44, Jun. 2015.
[50] A. Ghose and P. G. Ipeirotis, “Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics,” *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 10, pp. 1498–1512, Oct. 2011.
[51] N. Korfiatis, E. García-Bariocanal, and S. Sánchez-Alonso, “Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content,” *Electron. Commer. Res. Appl.*, vol. 11, no. 3, pp. 205–217, May 2012.
[52] L. M. Willemsen, P. C. Neijens, F. Bronner, and J. A. de Ridder, “‘Highly Recommended!’ The Content Characteristics and Perceived Usefulness of Online Consumer Reviews,” *J. Comput.-Mediat. Commun.*, vol. 17, no. 1, pp. 19–38, Oct. 2011.
[53] S. Park and J. L. Nicolau, “Asymmetric effects of online consumer reviews,” *Ann. Tour. Res.*, vol. 50, no. Supplement C, pp. 67–83, Jan. 2015.
[54] J. Liu, Y. Cao, C. Lin, Y. Huang, and M. Zhou, “Low-Quality Product Review Detection in Opinion Summarization,” presented at the Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), 2007, pp. 334–342.
[55] M. P. O’Mahony and B. Smyth, “Learning to Recommend Helpful Hotel Reviews,” in *Proceedings of the Third ACM Conference on Recommender Systems*, New York, NY, USA, 2009, pp. 305–308.
[56] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, Stroudsburg, PA, USA, 2002, pp. 79–86.
[57] P. Turney, “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews,” 2002. [Online]. Available: http://arxiv.org/abs/cs.LG/0212032. [Accessed: 01-Nov-2011].

[58] B. Pang and L. Lee, “A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts,” in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Stroudsburg, PA, USA, 2004.
[59] S. Kim and E. Hovy, “Determining the sentiment of opinions,” in *In Proceedings of the 20th International Conference on Computational Linguistics (COLING’04*, 2004, pp. 1367–1373.
[60] E. Riloff and J. Wiebe, “Learning extraction patterns for subjective expressions,” in *Conference on Empirical Methods in Natural Language Processing (EMNLP-03*, 2003, pp. 105–112.
[61] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, “Lexicon-Based Methods for Sentiment Analysis,” *Comput. Linguist.*, vol. 37, no. 2, pp. 267–307, Apr. 2011.
[62] T. Wilson *et al.*, “OpinionFinder: A System for Subjectivity Analysis.,” in *HLT/EMNLP 2005: 2005*, 2005.
[63] M. Hu and B. Liu, “Mining and summarizing customer reviews,” presented at the KDD, 2004, pp. 168–177.
[64] S. Blair-Goldensohn, K. Hannan, and R. McDonald, “Building a Sentiment Summarizer for Local Service Reviews,” presented at the NLPIX, 2008.
[65] H. Wang, Y. Lu, and C. Zhai, “Latent aspect rating analysis on review text data: a rating regression approach ‘KDD’, pp. 783-792,” presented at the KDD, 2010, pp. 783–792.
[66] T. Hofmann, “Probabilistic latent semantic indexing,” presented at the SIGIR, 1999, pp. 50--57.
[67] M. B. David, Y. N. Andrew, and I. J. Michael, “Latent Dirichlet Allocation,” *J. Mach. Learn. Res.*, vol. 3, pp. 933–1022, 2003.
[68] A. Gruber, M. Rosen-zvi, and Y. Weiss, “Hidden topic Markov models,” *Proc. Artif. Intell. Stat.*, 2007.
[69] I. Titov and R. McDonald, “Modeling online reviews with multi-grain topic models,” 2008, pp. 111–120.
[70] S. Moghaddam and M. Ester, “ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews,” in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information*, New York, NY, USA, 2011, pp. 665–674.
[71] C. Lin and Y. He, “Joint sentiment/topic model for sentiment analysis,” in *Proceeding of the 18th ACM conference on Information and knowledge management*, New York, NY, USA, 2009, pp. 375–384.
[72] Y. Jo and A. Oh, “Aspect and sentiment unification model for online review analysis,” presented at the WSDM, 2011, pp. 815–824.
[73] Y. Lu, C. Zhai, and N. Sundaresan, “Rated aspect summarization of short comments,” in *Proceedings of the 18th international conference on World wide web*, New York, NY, USA, 2009, pp. 131–140.
[74] D. M. Blei and J. D. Lafferty, “Correlated topic models,” *Proc. 23RD Int. Conf. Mach. Learn.*, pp. 113--120, 2006.

[75] D. M. Blei and J. D. Lafferty, “Dynamic topic models,” in *Proceedings of the 23rd international conference on Machine learning*, New York, NY, USA, 2006, pp. 113–120.
[76] W. Li and A. McCallum, “Pachinko allocation: DAG-structured mixture models of topic correlations,” in *Proceedings of the 23rd international conference on Machine learning*, New York, NY, USA, 2006, pp. 577–584.
[77] T. M. Department, T. Minka, and J. Lafferty, “Expectation-Propagation for the Generative Aspect Model,” *Proc. 18TH Conf. Uncertain. Artif. Intell.*, pp. 352--359, 2002.
[78] T. L. Griffiths, “Finding scientific topics,” *Proc. Natl. Acad. Sci.*, vol. 101, pp. 5228–5235, Jan. 2004.
[79] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, “Introducing markov chain monte carlo,” in *Markov chain Monte Carlo in practice*, Springer, 1996, pp. 1–19.
[80] C. C. Clogg, “Latent Class Models,” in *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, Springer, Boston, MA, 1995, pp. 311–359.
[81] L. M. Collins and S. T. Lanza, “Latent Class Analysis with Covariates,” in *Latent Class and Latent Transition Analysis*, John Wiley & Sons, Inc., 2009, pp. 149–177.
[82] P. F. Lazarsfeld and N. W. Henry, *Latent structure analysis*. Houghton, Mifflin, 1968.
[83] L. A. Goodman, “Exploratory latent structure analysis using both identifiable and unidentifiable models,” *Biometrika*, vol. 61, no. 2, pp. 215–231, Aug. 1974.
[84] J. Cha, “Application of Ordered Latent Class Regression in Educational Assessment,” Columbia University, 2011.
[85] J. A. Hagenaars, *Loglinear models with latent variables*, vol. iv. Thousand Oaks, CA, US: Sage Publications, Inc, 1993.
[86] K. Yamaguchi, “Multinomial Logit Latent-Class Regression Models: An Analysis of the Predictors of Gender Role Attitudes among Japanese Women,” *Am. J. Sociol.*, vol. 105, no. 6, pp. 1702–1740, May 2000.
[87] A. K. W. R. Dillon, “Latent structure and other mixture models in marketing: an integrative survey and overview.”
[88] *Applications of latent trait and latent class models in the social sciences*. New York, NY, US: Waxmann Publishing Co, 1997.
[89] W. W. Eaton, A. Dryman, A. Sorenson, and A. McCutcheon, “DSM-III major depressive disorder in the community. A latent class analysis of data from the NIMH epidemiologic catchment area programme.,” *Br. J. Psychiatry*, vol. 155, no. 1, pp. 48–54, Jul. 1989.
[90] P. S. Albert, L. M. McShane, J. H. Shih, and T. U. S. N. C. I. B. T. M. Network, “Latent Class Modeling Approaches for Assessing Diagnostic Error without a Gold Standard: With Applications to p53 Immunohistochemical Assays in Bladder Tumors,” *Biometrics*, vol. 57, no. 2, pp. 610–619, 2001.
[91] A. K. Formann and T. Kohlmann, “Latent class analysis in medical research,” *Stat. Methods Med. Res.*, vol. 5, no. 2, pp. 179–211, Jun. 1996.
[92] E. S. Garrett and S. L. Zeger, “Latent Class Model Diagnosis,” *Biometrics*, vol. 56, no. 4, pp. 1055–1067, 2000.

[93] R. J. Neuman *et al.*, “Latent Class Analysis of ADHD and Comorbid Symptoms in a Population Sample of Adolescent Female Twins,” *J. Child Psychol. Psychiatry*, vol. 42, no. 7, pp. 933–942, 2001.
[94] J. Magidson and J. Vermunt, “Latent Class Models for Clustering: A Comparison with K-means,” *Can J Mark. Res.*, vol. 20, Nov. 2001.
[95] J. B. Schreiber and A. J. Pekarik, “Technical Note: Using Latent Class Analysis versus K-means or Hierarchical Clustering to Understand Museum Visitors,” *Curator Mus. J.*, vol. 57, no. 1, pp. 45–59, Jan. 2014.
[96] P. Kent, R. K. Jensen, and A. Kongsted, “A comparison of three clustering methods for finding subgroups in MRI, SMS or clinical data: SPSS TwoStep Cluster analysis, Latent Gold and SNOB,” *BMC Med. Res. Methodol.*, vol. 14, p. 113, Oct. 2014.
[97] M. J. Green, “Latent class analysis was accurate but sensitive in data simulations,” *J. Clin. Epidemiol.*, vol. 67, no. 10, pp. 1157–1162, Oct. 2014.
[98] J. K. Vermunt and J. Magidson, “Factor analysis with categorical indicators: A comparison between traditional and latent class approaches,” *New Dev. Categ. Data Anal. Soc. Behav. Sci.*, pp. 41–62, 2005.
[99] S. Lee and J. Y. Choeh, “Predicting the Helpfulness of Online Reviews Using Multilayer Perceptron Neural Networks,” *Expert Syst Appl*, vol. 41, no. 6, pp. 3041–3046, May 2014.
[100] L. T. Su, H. Chen, and X. Dong, “Evaluation of Web-Based Search Engines from the End-User’s Perspective: A Pilot Study,” *Proc. ASIS Annu. Meet.*, vol. 35, pp. 348–61, 1998.
[101] D. Hawking, N. Craswell, P. Bailey, and K. Griffihs, “Measuring Search Engine Quality,” *Inf. Retr.*, vol. 4, no. 1, pp. 33–59, Apr. 2001.
[102] L. Vaughan, “New measurements for search engine evaluation proposed and tested,” *Inf. Process. Manag.*, vol. 40, no. 4, pp. 677–691, Jul. 2004.
[103] M. M. S. Beg, “A subjective measure of web search quality,” *Inf. Sci.*, vol. 169, no. 3, pp. 365–381, Feb. 2005.
[104] D. Kelly, “Methods for Evaluating Interactive Information Retrieval Systems with Users,” *Found. Trends® Inf. Retr.*, vol. 3, no. 1—2, pp. 1–224, 2007.
[105] T. L. Saaty, “What is the Analytic Hierarchy Process?,” in *Mathematical Models for Decision Support*, Springer, Berlin, Heidelberg, 1988, pp. 109–121.
[106] M. G. Kendall, “A NEW MEASURE OF RANK CORRELATION,” *Biometrika*, vol. 30, no. 1–2, pp. 81–93, Jun. 1938.
[107] T. L. Saaty, *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York ; London: Mcgraw-Hill, 1980.
[108] L. L. Thurstone, “A Law of Comparative Judgment,” *Psychol. Rev.*, vol. 34, no. 4, pp. 273–286, 1927.
[109] T. L. Saaty, “A scaling method for priorities in hierarchical structures,” *J. Math. Psychol.*, vol. 15, no. 3, pp. 234–281, Jun. 1977.
[110] F. A. Lootsma, “Scale sensitivity in the multiplicative AHP and SMART,” *J. Multi-Criteria Decis. Anal.*, vol. 2, no. 2, pp. 87–110, Aug. 1993.
[111] A. A. Salo and R. P. Hämäläinen, “On the measurement of preferences in the analytic hierarchy process,” *J. Multi-Criteria Decis. Anal.*, vol. 6, no. 6, pp. 309–319, Nov. 1997.

[112] Y. Wind and T. L. Saaty, “Marketing Applications of the Analytic Hierarchy Process,” *Manag. Sci.*, vol. 26, no. 7, pp. 641–658, Jul. 1980.
[113] M. Davies, “Adaptive AHP: a review of marketing applications with extensions,” *Eur. J. Mark.*, vol. 35, no. 7/8, pp. 872–894, Aug. 2001.
[114] S. D. Pohekar and M. Ramachandran, “Application of multi-criteria decision making to sustainable energy planning—A review,” *Renew. Sustain. Energy Rev.*, vol. 8, no. 4, pp. 365–381, Aug. 2004.
[115] C. Kahraman, İ. Kaya, and S. Cebi, “A comparative analysis for multiattribute selection among renewable energy alternatives using fuzzy axiomatic design and fuzzy analytic hierarchy process,” *Energy*, vol. 34, no. 10, pp. 1603–1616, Oct. 2009.
[116] M. J. Liberatore and R. L. Nydick, “The analytic hierarchy process in medical and health care decision making: A literature review,” *Eur. J. Oper. Res.*, vol. 189, no. 1, pp. 194–207, Aug. 2008.
[117] C. W. Lee, N. K. Kwak, and Correspondence, “Information resource planning for a health-care system using an AHP-based goal programming method,” *J. Oper. Res. Soc.*, vol. 50, no. 12, pp. 1191–1198, Dec. 1999.
[118] K. Heidenberger and C. Stummer, “Research and development project selection and resource allocation: a review of quantitative modelling approaches,” *Int. J. Manag. Rev.*, vol. 1, no. 2, pp. 197–224, Jun. 1999.
[119] A. N. Langville and C. D. Meyer, “A Survey of Eigenvector Methods for Web Information Retrieval,” *SIAM Rev.*, vol. 47, no. 1, pp. 135–161, Mar. 2005.
[120] R. Fagin, R. Kumar, and D. Sivakumar, “Comparing Top K Lists,” in *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, Philadelphia, PA, USA, 2003, pp. 28–36.
[121] M. F. Porter, “An algorithm for suffix stripping,” *Program Electron. Libr. Inf. Syst.*, vol. 14, no. 3, pp. 130–137, 1980.
[122] W. Buntine, “Estimating Likelihoods for Topic Models,” in *Advances in Machine Learning*, 2009, pp. 51–64.
[123] I. Murray and R. R. Salakhutdinov, “Evaluating probabilities under high-dimensional latent variable models,” in *Advances in Neural Information Processing Systems 21*, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, Eds. Curran Associates, Inc., 2009, pp. 1137–1144.
[124] J. Chang and D. M. Blei, “Relational Topic Models for Document Networks.,” *J. Mach. Learn. Res. - Proc. Track*, vol. 5, pp. 81–88, Jan. 2009.
[125] D. M Blei, “Introduction to Probabilistic Topic Models,” *Commun. ACM*, vol. 55, Jan. 2011.
[126] “Reading Tea Leaves: How Humans Interpret Topic Models.” [Online]. Available: http://videolectures.net/nips09_boyd_graber_rtl/. [Accessed: 02-Jan-2018].
[127] Q. Mei, C. Liu, H. Su, and C. Zhai, “A Probabilistic Approach to Spatiotemporal Theme Pattern Mining on Weblogs,” in *Proceedings of the 15th International Conference on World Wide Web*, New York, NY, USA, 2006, pp. 533–542.
[128] X. Wang and A. McCallum, “Topics over time: a non-Markov continuous-time model of topical trends,” in *Proceedings of the 12th ACM SIGKDD*

international conference on Knowledge discovery and data mining, New York, NY, USA, 2006, pp. 424–433.
[129] Z. Li, J. Feng, and Z. Xiao-Yan, "Movie review mining and summarization," presented at the CIKM '06 Proceedings of the 15th ACM international conference on Information and knowledge management, 2006.
[130] B. M. Byrne, "Structural Equation Modeling: Perspectives on the Present and the Future," *Int. J. Test.*, vol. 1, no. 3–4, pp. 327–334, Sep. 2001.
[131] P. F. Wu, V. D. Heijden, Hans, and N. Korfiatis, "The Influences of Negativity and Review Quality on the Helpfulness of Online Reviews," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 1937664, Aug. 2011.
[132] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky, "The Stanford CoreNLP Natural Language Processing Toolkit," in *Association for Computational Linguistics (ACL) System Demonstrations*, 2014, pp. 55–60.
[133] F. Å. Nielsen, "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs," presented at the ESWC2011 Workshop on "Making Sense of Microposts": Big things come in small packages, Heraklion, Crete, 2011, pp. 93–98.
[134] W. Hong, J. Y. L. Thong, W.-M. Wong, and K.-Y. Tam, "Determinants of User Acceptance of Digital Libraries: An Empirical Examination of Individual Differences and System Characteristics," *J. Manag. Inf. Syst.*, vol. 18, no. 3, pp. 97–124, Jan. 2002.
[135] D. Riecken, "Personalized Views of PERSONALIZATION," *Commun. ACM*, vol. 43, no. 8, pp. 26–26, Aug. 2000.
[136] K. Y. Tam and S. Y. Ho, "Understanding the Impact of Web Personalization on User Information Processing and Decision Outcomes," *MIS Q.*, vol. 30, no. 4, pp. 865–890, 2006.
[137] J. G. Maxham and R. G. Netemeyer, "A Longitudinal Study of Complaining Customers' Evaluations of Multiple Service Failures and Recovery Efforts," *J. Mark.*, vol. 66, no. 4, pp. 57–71, Oct. 2002.
[138] A. Mantonakis, P. Rodero, I. Lesschaeve, and R. Hastie, "Order in Choice: Effects of Serial Position on Preferences," *Psychol. Sci.*, vol. 20, no. 11, pp. 1309–1312, Nov. 2009.
[139] B. Ives, M. H. Olson, and J. J. Baroudi, "The Measurement of User Information Satisfaction," *Commun ACM*, vol. 26, no. 10, pp. 785–793, Oct. 1983.
[140] W. Delone and E. McLean, "Information Systems Success: The Quest for the Dependent Variable," *Inf. Syst. Res.*, vol. 3, pp. 60–95, Mar. 1992.