Impact of personalized review summaries on buying decisions: An experimental study

Mahesh Balan
Department of Management Studies,
Indian Institute of Technology Madras
u.maheshbalan@gmail.com

Saji K Mathew
Department of Management Studies,
Indian Institute of Technology Madras
saji@iitm.ac.in

Abstract
This study evaluates the impact of personalization of review summaries on consumers’ cognitive efforts and buying decision. Following an experimental procedure we tested four hypotheses pertaining to online buyers’ decision process. Our results show that personalized review summary significantly reduces the information processing effort and information requirements of those who received personalized review summaries as compared to those who did not. This study thus contributes to e-commerce literature on online buyer behavior and recommender systems strategy.

Keywords
Consumer reviews, review summary, text-mining, personalization, cognitive processing

Introduction
With the steady growth of e-commerce triggering the growth of consumer segment who shop online, online word of mouth and its influence on buying decisions have been the focus of recent research in information systems (Maeyer, 2012; Pang and Lee, 2008). Extending the definition of word of mouth (WOM) given by Arndt (1967), consumer product reviews or online word of mouth can be defined as a message from the person who bought the product online about her or his experience to a person who is interested in buying a product, where the person who is giving the review message has a completely non-commercial intention. The primary sources of online user generated contents are consumer product ratings and consumer product reviews. Prior studies have empirically shown that online reviews significantly affect product sales (Chevalier and Mayzlin, 2006; Liu et al., 2006; Ghose and Ipeirotis, 2011). It is also found that consumer product ratings which are numerical in nature follows a bimodal distribution (Hu et al. 2009) and consumer product reviews give a bigger picture about the product's than numerical product ratings by giving a 360 degree overview of the product experience (Archak et al., 2011).

Another stream of research in e-commerce has investigated the impact of web personalization in customers’ information processing and decision outcomes (Tam and Ho, 2006; Ho et al., 2010; Vinodh et al., 2015). Web personalization also covers content and presentation formats tailored to user preferences. Personalization has been studied widely in all domains, for example Google news has been personalized based on user behavior providing relevant items that are beneficial to the user (Das et al., 2007; Mobasher et al., 2000). However the impact of personalization on consumers’ decisions for different categories of content in e-commerce remains under explored. Also prior research in this area has concentrated more upon the effect of reviews on sales, and the dimensionality of reviews but very less research has focused on the behavior of the consumer and the impact of reviews on the consumer's purchase decision and thought process.

This paper studies the effect of personalized review summaries on consumer’s buying decision. We study the effect of personalized review summarization on consumer’s buying decision and how it affects their cognitive process. It can be considered as an extension work by Archak et al. (2011) where they evaluated the pricing power of consumer reviews and also by Tam and Ho (2006) where they evaluated the effect of personalization on consumer’s cognitive process while making a buying decision. This study involves two parts – First, text mining of reviews, tagging them by the features they are talking about and finding the
sentiment score for the review. Second, providing a summary of the review sentiment scores for every feature to the user, thus making it easier for the user to get the summary of reviews. The study also involved a multi-factor multi-group experiment for validation where a virtual e-commerce mobile shop was created. The consumer’s buying decision in different treatments like the e-commerce site with personalized review summary and non-personalized review summary and also in goal specific and non-goal specific treatments were studied in detail. The study shows strong evidence that personalized review summarization affects consumer’s buying decision.

Our study makes two important contributions to the body of knowledge related to Internet commerce. First we study the impact of feature level summary of consumer reviews on consumer buying behavior. Feature extraction from voluminous text is expected to ease information processing and decision making behavior of consumers. Our work is thus very close to the study of Archak et al. (2011), where they evaluate the weights given by consumers on every product feature from the consumer reviews. Our paper uses concepts from their work for extracting product features from reviews. Second, we further personalize feature summaries using the stated preferences of consumers. This is expected to further ease the cognitive effort of users in information processing and is expedited to impact buying behavior. This paper also uses literature from personalization, social cognition and psychology. This work can be considered as an extension of the prior works of Archak et al. (2011), and Tam and Ho (2006), who evaluated the impact of personalization on consumer’s decision making, thus contributing to information systems literature on understanding impact of consumer review systems on consumer’s buying decision.

**Literature Review**

Research in online word of mouth addresses many different research gaps like studying the effect of reviews on sales, reviewer behavior and bias, dimensionality of reviews and economic impact of reviews, also termed as economining. Sources of online word of mouth in e-commerce platforms can be both numerical data in the form of consumer product ratings and textual data in the form of consumer product reviews. There is significant evidence from prior literature that online word of mouth affects product sales (Chevalier and Mayzlin, 2006; Liu et al., 2006; Ghose and Ipeirotis, 2011; Archak et al., 2011), but most of them provide evidence only by using numerical data which is the consumer product ratings.

**Online Product Reviews**

Several studies have analyzed the effect of consumer reviews in sales performance in various segments like books (Chevalier and Mayzlin, 2006), movies (Liu et al., 2006) and professional services (Gao et al., 2015). The seminal study of Chevalier and Mayzlin (2006) provided early empirical evidence that there is a significant relationship between online word of mouth and product sales. They showed that an increase in online reviews increased book sales in Amazon.com and Barnesandnoble.com. They also showed the vice-versa is true. Like other previous studies, this study also used Amazon seller ranks as a proxy for product demand. Chevalier and Mayzlin (2006) found three important reasons for ex-ante effect in consumer reviews. First, reviewers are not incentivized and hence no motivation to write a review. Second, switching costs are low and users use review information from one website to buy products in the other. Third, the e-commerce platform has control to edit information being posted there. The positive quartile of reviews were playing a significant role in product sales. The same significant relationship between online word of mouth and sales was also observed in the entertainment industry. Liu et al. (2006) studied Yahoo! Movie reviews and found that the effect of online word of mouth on box-office performance is more during one week pre-release and one week post-release duration.

Recently research has also focused on the reviewer behavior and bias. Though prior work has suggested that product recommendations increase the probability of the buyer to choose the product (Senecal and Nantel, 2004), there are many other factors that affect the behavior of the buyer before she or he chooses the product, like helpfulness of a review (Ghose and Ipeirotis, 2011) and seller reputation (Resnick et al., 2006). Fradkin et al. (2015) studied reviewer behavior by conducting three field experiments in aribnb.com, a large online accommodation provider platform. They found that non-reviewers tend to have bad experiences than reviewers. The reviews are biased when asked for reviewer information and when the same reviewer is asked to comment anonymously, the opinions are different from that in the original review. Scholars have used techniques from computer science like Natural Language Processing (NLP) and sentiment analysis to study the dimensionality of reviews. Kim et al. (2004) used reviews of MP3 players and cameras to study the dimensionality of online word of mouth. They broke down every review...
into various components like structure (word length, and frequency of sentences), term document matrix (lexical), meta data features (product rating) and found that the useful characteristics that can be used to rank the reviews are review helpfulness, review length, lexical features like morphology and syntax, and rating given by the reviewer.

Feature summary techniques of consumer reviews have also been addressed in prior literature which involve text summarization (Zhuang et al., 2006), sentiment summarization (Beineke et al., 2004) and aspect level summary from reviews (Blair-Goldensohn et al., 2008). Many techniques from computer science like sentiment analysis, part-of-speech tagging and named entity recognition are applied to consumer reviews in e-commerce to extract meaning from them. Consumer reviews are unstructured in nature, huge in volume, serial in presentation and hence consumers underutilize them due to their limited search capabilities. This brings a need for a visualization summary for consumers like Opinion Blocks (Alper et al., 2011). Blair-Goldensohn et al. (2008) extracted sentiment summary for relevant aspects of a restaurant like food, ambience, service and value. They tagged the reviews by these aspects and summarized their sentiments. In order to eliminate over-exaggerated reviews subjectivity score was also computed for the reviews.

**Web Personalization**

Web personalization refers to the process of adapting web content to meet the specific needs of users to maximize sales (Korper and Ellis, 2001). Tam and Ho (2006) studied the impact of web personalization on the decision making process of consumers. They conducted a multi-group multi-factor experiment to test the effect of self-reference, content relevance and goal specificity on the cognitive process of consumers while making a buying decision. This study reported that self-reference, content relevance and goal specificity influence cognitive processes, attention and purchase decisions of consumers. Further Slovic (1995) showed that customers with the same preference processes web information, but the preferences elicited through these processes need not be the same for every person. Consumers find it really difficult to make their choices when there are so many choices available (Iyengar et al., 2000). So, they look for more information to justify that their choice is rational (Yaniv et al., 2004). The additional information can be in the form of product ratings and product reviews.

Drawing on the foundations of social cognition research, prior studies have conceptualized web personalization as consisting of two important constructs: self-reference and content relevance (Wyer and Srull, 1989; Tam and Ho, 2006). The cognitive process is the set of steps the consumer engages inside his or her brain during his thought process before making a purchase decision for a particular product. These cognitive processes are dependent upon other factors such as content relevance in the website and also self-reference. Self-reference is the ability to process information related to the self distinctly from processing other information. The famous example is the cocktail-party example, where an individual has the ability to detect her or his name being called despite the noise inside the party hall. The agents that personalize contents for a user are called personalizing agents and most of them personalize on the basis of content relevance and self-reference only.

Although recent research has been actively investigating the role of consumer reviews in e-commerce, most of the prior work addresses only reviewer behavior and hardly few studies deal with the effect of consumer reviews on consumer’s purchase behavior. Personalized feature summaries have also been not addressed in prior studies to the best of our knowledge. The purpose of our study is to understand the effect of personalized features summaries extracted from consumer reviews on online buying behavior. By providing personalized feature level summary sentiment scores of every product to the user, we test both the consumer’s buying decision and also the cognitive efforts put by them to make the buying decision.

**Hypotheses Development**

Social cognition is the study of construction of social reality of a user which is a result of the user’s cognitive process (Wyer and Srull, 1989). Choice made by a person is the result of information processing in the brain (Huber and Seiser, 2001). The human brain has a permanent memory and working memory. The permanent memory stores all information and it has long time memory, whereas the working memory provides temporary storage and keeps all information that are relevant for the current scenario. When the consumer visits a website to buy a product with some prior belief and all the information related to this belief will be stored in the working memory. A personalized review summary can help the
user find information to update their beliefs (Archak et al., 2011, p. 1487) easily by putting lesser cognitive efforts. This ease in cognitive effort will be reflected in the number of items they add to the cart (and later remove) for a purchase activity.

\[ H1: \text{Users when exposed to personalized review summary will add lesser number of items to the cart to buy a product as compared to those who do not receive personalized review summary} \]

Every piece of information given to the users in the website is a form of stimulus and it can be in the form of text or image. These stimuli grab the user’s attention and induce them to explore more about a product. One measure that is a good proxy to explain this behavior of the user is the number of clicks made by a user in a website during a session (Tam and Ho, 2006; Archak et al., 2011). The users click the product to seek more information about the product like detailed product description and product reviews. This need to explore more information (through clicks) is hypothesized to be lesser in an e-commerce environment with a personalized review summary.

\[ H2: \text{Users when exposed to personalized review summary will seek lesser information about the product to make their buying decision as compared to those who do not receive personalized review summary} \]

The feature summary was provided in a personalized manner, but its behavior in a non-personalized environment is not known. In addition to the non-personalized no review summary website, we added one more website that has a feature based summary of reviews in a non-personalized manner. This will make the different levels of personalization to three. Based on these, we further hypothesize the following:

\[ H3: \text{Users when exposed to personalized review summary will add lesser number of items to the cart to buy a product as compared to those who receive a non-personalized review summary and also those who do not receive review summary} \]

\[ H4: \text{Users when exposed to personalized review summary will seek lesser information about the product to make their buying decision as compared to those who receive a non-personalized review summary and also those who do not receive review summary} \]

**Methodology**

Our study follows an empirical approach employing a quasi-experimental procedure. The study is broken down into two parts – first, text mining of reviews to extract product features and computing sentiment summary scores. Second, a factorial design of experiment to evaluate the impact of personalized review summary on consumer’s buying decision. Information provided in the e-commerce website can act as a stimulus that attracts consumer attention. We evaluate the effect of these stimuli on the consumer’s purchase decision through this experiment.

**Extracting Product Feature Sentiment Summary**

Consumers have a belief about the product before visiting the e-commerce platform. The belief follows a normal distribution from 0 to 1 with mean of 0.5 and it gets updated after reading the reviews. The belief distribution could get updated to a mean of 0.75 and as the consumer sees new information (Archak et al., 2011). Many techniques are available for extracting product features. We use part-of-speech tagging as it is widely used in prior research for feature extraction. Mobile phone and Tablet reviews were taken from Flipkart.com for 40 products comprising of 28 mobile phones and 12 tablets. We chose these products as they have clearly identifiable features and are familiar to college students. After applying part of speech tagging to the reviews, the nouns and noun phrases were eliminated. The top 20% of the nouns and noun phrases were taken as the product features of the mobile phone. The manufacture defined features were also added to this set of features extracted from POS tagging. So, the feature was a mix of both user defined and manufacture defined features. Manufacture defined features include camera, battery, price, sound and the like, whereas user defined features include panorama, selfie, music and the like. These set of features would then be used as reference to tag every review by the list of features it is talking about.
After the tagging, the sentiment score is computed for the review. But, one review can talk positive about one feature and negative about another feature. To overcome this challenge, sentiment was computed at the feature level of reviews that refer to more than one feature. For reviews that had more than one feature tag, a phrase was formed by joining three words to the right of the feature word and three words to the left of the feature word (Narayanan et al., 2009). A score was computed for this phrase formed, which gave the sentiment score for that particular feature. This eliminates the risk of false positives and false negatives in the sentiment polarity estimation of the features. The sentiment score ranges from -1 to +1. The average of the sentiment scores for a particular feature is the summary sentiment score for that particular feature. Thus, summary sentiment score was calculated for all the extracted product features. A summary chart comprising of data bars that shows feature level average sentiment scores was displayed along with the product image to give the consumers a better understanding of the product.

**Personalization of Review Summaries**

We used conjoint analysis for personalization of review summaries. Conjoint analysis is widely used in marketing research and also in other domains to analyze consumer’s stated preferences (Green and Srinivasan, 1978; Pang and Lee, 2008). There are two ways to elicit consumer preferences – stated preferences and revealed preferences. Stated preferences are the preferences explicitly stated by the consumer. Revealed preferences are those preferences that are extracted from the consumer without his/her knowledge through browsing history, purchase history and the like. In this paper, we use stated preferences and apply conjoint analysis on them to personalize the products based on review summaries. The products were ranked in the order of stated consumer preferences on review summaries as follows.

Top 5% of the features were extracted using part of speech tagging of consumer reviews which resulted in three features- battery, camera and price. Each of these three factors had two levels each – high sentiment and low sentiment. These levels describe the sentiment score of the respective features which are proxies for opinions coming from reviews for the respective features. The consumers ranked 8 combinations of the product features with their levels in the order of their preferences. Utility scores were calculated using conjoint analysis for each product which were specific to individual consumers based on their stated preferences. The products are then ranked in descending order of utility scores, thus personalizing the webpage. This personalization is based on the feature level average sentiment scores of the products and conjoint analysis converts the consumer stimuli to utility scores. By listing products in the order of consumer’s preference, we formed content relevance in personalization (Tam and Ho, 2006) for consumer stimuli in online environment.

**Experimental Procedure**

Three different versions of virtual e-commerce websites were developed exclusively for this experiment. All the three websites had the same user interface and same 40 products comprising 28 mobiles and 12 tablets. The mobile phones and tablets in all the websites had consumer reviews with helpfulness score, product ratings and number of people who voted the rating and manufacturer defined product description. Three different versions of web interfaces were designed: (i) one with reviews in original textual form (ii) second with reviews in textual form plus a summary of reviews as feature scores (iii) Third textual review plus personalized feature scores. The consumers have to sign up and create an account and they have to fill a form asking for basic demographical data and in the personalized setting they need to rank 8 products.

The experiment was conducted in the IT Laboratory of a premier engineering institute in Southern India. The sample population dominantly consisted of students from the same premier institute. The experiment was conducted in multiple batches with batch size ranging from 30 to 60 participants per batch. Our experiment followed a random sampling to allocate students to sub groups with three levels of personalization. For this purpose we used lottery method, by asking them to pick numbers from a lot. Before the start of the experiment, their working memory was cleared by playing fun videos and having other entertaining activities for some time. It was made sure that all the participants started the experiment at the same time. The participants were assigned a goal of buying 2 products mandatorily which could be mobiles or tablets. After the experiment, participants were asked to checkout and fill a survey instrument which was derived from previously tested and validated instruments. The questions related to personalization were taken from Tam and Ho (2006).
Impact of personalized review summaries on buying decisions

Analyses

Our sample consisted of 213 graduate students and after cleaning the data by dropping incomplete data and outliers, 195 cases were left. We used click stream data from server (MY SQL database) to extract cart updates and number of clicks by each user. The cart updates data is a measure of the total number of product additions and deletions from the user’s cart. Cart updates serves as a proxy for cognitive effort for buying decision. The clicks data is a measure of the number of clicks on the product details and product reviews.

To test our first hypothesis, we ran an independent sample t-test between cart updates and personalization (2 levels) for personalized and non-personalized review summary groups. The mean of the cart updates for non-personalized group was significantly higher than the cart updates for personalized review summary group (p = 0.032). The F statistic to evaluate homogeneity of variance between the two groups was also very significant, thus showing there is significant difference between the variances of cart updates of the two groups (F = 22.650, p < 0.001). These results support our first hypothesis that users exposed to personalized review summary will add lesser number of products to the cart to buy a product as compared to those who do not receive personalized review summary.

Another independent sample t-test was conducted to test our second hypothesis. We analyzed clicks for treatment and control groups with respect to feature summary based personalization. The mean clicks for non-personalized group was significantly higher than mean clicks for personalized review summary group (7.80 and 3.65, p = 0.042). The F statistic was significant, thus indicating difference in variances between clicks of the two groups (F = 17.857, p < 0.001).

<table>
<thead>
<tr>
<th>Measurement</th>
<th>F</th>
<th>Sig</th>
<th>Mean difference</th>
<th>Std error</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cart Updates</td>
<td>22.650</td>
<td>0.000</td>
<td>0.557</td>
<td>0.252</td>
<td>p = 0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(H1 supported)</td>
</tr>
<tr>
<td>Clicks</td>
<td>17.857</td>
<td>0.000</td>
<td>4.148</td>
<td>1.983</td>
<td>p = 0.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(H2 supported)</td>
</tr>
</tbody>
</table>

Table 1: Independent sample t-test for feature summary based personalization
This supports the second hypothesis which posited that users seek less information in personalized group than non-personalized. The lesser mean clicks for personalized group is a good measure of the user seeking lesser information about the product.

To identify other potential effects, we ran a one way ANOVA with the dependent variable being cart updates and the factor being personalization level (3 levels). The personalization has 3 levels here which are users who did not receive review summary, who received non-personalized review summary and who received personalized review summary. The mean of cart updates was found to be the lowest for users who received personalized review summary compared to those who received non-personalized review summary and those who did not receive review summary. The F statistic was also significant, hence indicating that there are significant differences in variances between the users in the three levels of personalization. The mean of cart updates was the highest for people who received non-personalized review summary and the reason for it is unknown and it will be explored further in future work. Thus, it can be inferred that personalized review summaries plays a significant role in cart updates, adding support to the third hypothesis.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Levels</th>
<th>Mean</th>
<th>F</th>
<th>Sig.</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cart Updates</td>
<td>No review summary</td>
<td>2.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-personalized review summary</td>
<td>2.60</td>
<td>3.102</td>
<td>0.05</td>
<td>H3 supported</td>
</tr>
<tr>
<td></td>
<td>Personalized review summary</td>
<td>2.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: ANOVA results of cart updates for 3 personalization levels

<table>
<thead>
<tr>
<th>Measure</th>
<th>Levels</th>
<th>Mean</th>
<th>F</th>
<th>Sig.</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clicks</td>
<td>No review summary</td>
<td>3.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-personalized review summary</td>
<td>7.80</td>
<td>3.503</td>
<td>0.036</td>
<td>H4 supported</td>
</tr>
<tr>
<td></td>
<td>Personalized review summary</td>
<td>3.65</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: ANOVA results of clicks for 3 personalization levels
Another one way ANOVA was applied with clicks being the dependent variable being clicks and the factor being personalization (3 levels) as explained above. The mean of clicks was found to be the lowest for users who received personalized review summary compared to those who received non-personalized review summary and those who did not receive review summary. The F statistic was also significant, hence indicating that there are significant differences in variances between the users in the three levels of personalization. The pattern in the data of clicks for the three levels of personalization is same as the pattern for updates data. The mean of clicks was the highest for people who received non-personalized review summary. Thus, it can be inferred that personalized review summaries makes users seek lesser information about the product, adding support to the fourth hypothesis.

Conclusion and Future Work

The theoretical contributions of this paper includes proposing a system of personalized review summaries for the consumers. We are the first to attempt to personalize products listing based on review summaries. The study also involves experimental validation of the same on the consumer’s buying decision and cognitive efforts needed to make the buying decision. The key findings show that when consumers are exposed to personalized review summary, they are less confused in making the buying decision and hence requiring lesser cognitive efforts in making the buying decision. The study also showed empirically that consumers seek lesser information about the product in an e-commerce environment with personalized review summary.

The results from our study have a few potential managerial implications. In an e-commerce environment, there is too much of information available, thus creating difficulty for the consumer in buying decision process. Famous e-commerce platforms have thousands of reviews and it is really difficult to read all the reviews and then take a decision. In our proposed new approach, the reviews are summarized and also the consumer will find it much easier to make a purchase decision. This is because a relevant brief summary minimizes the efforts to read reviews and extract relevant information. Apart from this, our study also benefits the e-commerce platform provider. The empirical results of this study show that consumers tend to click less in personalized review summary environment compared to non-personalized environment. Such information could be leveraged for personalization and recommender systems strategy.

The study has lots of scope for future work. First, the sample size can be increased and the behavior can be observed with many other sources of data other than cart updates and clicks. More complex statistical models can be built and the same behavior can be validated. Thus, this work motivates further research in this area by showing significant results.

References


