Exploring Hidden Factors behind Online Food Shopping from Amazon Reviews: A Topic Mining Approach

Yan Heng[a] , Yuan Jiang[b], and Zhifeng Gao[c]

[a]Post-doctoral Associate, Food and Resource Economics Department, PO Box 110240, University of Florida, Gainesville, Florida, 32611, yheng@ufl.edu

[b]Graduate Student, Food and Resource Economics Department, PO Box 110240, University of Florida, Gainesville, Florida, 32611, jyspring@ufl.edu

[c]Associate Professor, Food and Resource Economics Department, PO Box 110240, University of Florida, Gainesville, Florida, 32611, zfgao@ufl.edu

Selected Presentation at the Agricultural & Applied Economics Association's 2017 AAEA Annual Meeting, Chicago, IL, July 30-August 1, 2017

Introduction

Online sales have become an important channel for convenient shopping with a wide range of product selections. In the first quarter of 2017, the total U.S. retail e-commerce sales reached $105.7 billion, an increase of 4.1% from the fourth quarter of 2016 (U.S. Census, 2017). Online sales are expected to grow at an annual rate of 9.3% by 2020 (Forrester, 2016). Along with this trend, grocery online shopping is as well growing. According to a Nielson report, 25% of online respondents purchase grocery online, and over half are willing to do so in the future (Nielsen, 2015). As consumer's purchase decisions will affect sales and revenues directly, understanding how consumer make decisions during online shopping has become an essential subject for both researchers and e-commerce companies.

One influential factor affecting consumer online shopping decision-making is customer reviews (Duan et al., 2008; Chen et al. 2008). Online customers usually find themselves overwhelmed with a large number of competing products and overloaded information, yet they have limited time and knowledge to make effective decisions. Online retail websites, such as Amazon.com, provide consumers a platform to post product reviews to share their opinions and experience on products. These reviews, in turn, help potential consumers to make more effective decisions and attract more consumers (Cao et al. 2011). Although consumers would benefit from more information available, many products have hundreds and even thousands of reviews with various contents and inconsistent opinions, which makes it hard for consumers to use the information effectively for decision making. In this sense, online retailers highlight valuable information by the "helpfulness" feature, which allows consumers to evaluate other user's reviews. For example, Amazon.com asks "Was this review helpful to you?" under each review and makes this votes available alongside to shoppers. These websites also prioritize online customer reviews based on helpfulness votes to reduce consumers' cost of finding useful information. Moreover, customer reviews help online retailers develop positioning strategies because review text itself is a rich source of information on consumer's preferences and behavior (Moon and Kamaura, 2017).

Following the trend of big data usage, many studies have examined the factors that influence review helpfulness. For example, review rating (Cao et al, 2011; Mudambi and Schuff, 2010), the length of words (Kim et al., 2006; Hao et al., 2009), positive and negative sentiment (Berger et al. 2010), and recently emotions (Felbermayr and Nanopoulos, 2016) have been used to explain and predict the helpfulness votes. These studies mainly focus on improving review system design to encourage more helpfulness votes. However, there are no consistent conclusions regarding the important factors affecting the helpfulness of online reviews.

In this study, we hypothesize that those helpfulness votes reflect consumer's demand for some information that matters to them but hidden behind the review text itself. Consumers' evaluation of reviews depends on the textual information contained in reviews rather than solely on the numeric summary, such as length of words and review rating (Chevalier and Mayzlin, 2006). Therefore, ignoring text content of reviews is a major shortcoming of existing studies on recommender systems (McAuley and Leskovec, 2013). In other words, user reviews are text information with high dimensionality, and there are multiple latent interpretable topics underlying the texts and determining the helpfulness of the review. For example, Huang et al. (2013) showed that Yelp reviews on restaurants can be broken into some latent subtopics, including service, value, décor, and healthiness. In this sense, the high-dimensional review data can be processed by topic mining approaches to extract the low-dimensional latent topics.

Further differentiating from previous studies, we examined the factors affecting the helpfulness of customer reviews of food products that are rarely studied in this area. Products can be generally categorized into search goods and experience goods. A search good is a product whose quality is observable before purchase; and an experience good is a product whose quality can only be revealed after consuming or experiencing it (Nelson, 1970). Previous studies have examined customer reviews on search products, such as digital cameras (Lee and Bradlow, 2011; Archak et al., 2011), while some examined reviews on experience goods, including software programs (Cao et al., 2011) and sedan cars (Netzer et al., 2012). Mudamibi and Schuff (2010) compared customer reviews on research goods including digital cameras, cell phone, and laser printer, as well as experience goods including music CD, MP3 player, and video game. They found product type has an impact on the helpfulness of a review. All the aforementioned products have well-defined standards for product attributes and evaluations. Different from these products, food is a unique category of experience goods. Food consumers usually have more heterogeneous preferences and tastes, and it is not easy to consistently define important attributes associated with food products. The perceived quality of food can be highly subjective, and this is particularly true for sensorial products, such as wine and coffee (Moon and Kamaura, 2017). Till now, few studies have explored consumer reviews focusing on food products. Moon and Kamaura (2017) studied wine as an experience sensorial product in the literature, however, wine evaluation is more likely to rely on experts rather than consumers, which is different from most of the food products.

To fill the gap in the literature, our study aims to discover hidden topics behind online food consumer's review texts, and we further explore how these topics alone other review factors influence online shoppers. To achieve this goal, we apply the Latent Dirichlet Application

(LDA) factor model to extract latent subtropics from the customer reviews, which can help us justify the informative reviews and identify the information valued by customers. Additionally, to examine the effectiveness of the hidden factors in explaining the variation in helpfulness votes, we compare the interpretation power of hidden factors and numeric review attributes, including review rating and review length that are widely used in previous studies.

Our results show that four interpretable subtopics can be derived from the online reviews of coffee, and including subtopics can improve the interpretation and prediction of the helpfulness votes. Our study yields several interesting findings. First, we confirm that the review length has a positive but non-linear impact, while overall rating has a negative impact on helpfulness votes. Second, review readers perceive that underlying topics are helpful in assisting them to make online shopping decisions. Third, review readers value more the reviews that provide objective evidence, while reviews discussing subjective feeling and emotions are less likely to receive helpfulness votes.

Data

We use Amazon review data on grocery and food products from 2004 to 2014 provided by McAuley et al. (2015) in JSON file. There are total 1,048,576 distinct reviews in the dataset with information on reviewer ID, ASIN code, reviewer name, helpfulness vote, review text, overall rating, and review time. After examining the dataset, we selected coffee as the target product in this study for two major reasons. First, coffee products have the largest number of reviews among all food products. Second, coffee is a good example of sensorial food products with various differentiations. We finally selected 11 distinct coffee each having more than 1,000 reviews, which results in a total of 22,424 corresponding textual reviews.

For each review, we convert the numeric review summaries into variables: the total number of people who voted that the review was helpful (*Helpfulness*), the overall rating of the product (1 to 5) given by the reviewer (*Rating*), and the character count of the review (*Length*). We use the helpfulness votes as a dependent variable in the regression models in next section. The overall rating and review length are used as independent variables based on literature (Mudambi and Schuff, 2010; Kim et al., 2006; Hao et al., 2009). The overall rating and review length are factors that provide readers a quick impression of a review. Rating is also a brief identification of a consumer's evaluation and experience. The descriptive statistics are displayed in table 1. The average review on coffee products is positive with an average rating of 4.40. On average, each review has about 259 characters and 0.59 vote of helpfulness. The number of helpfulness votes varies from zero to 590 for one review.

The review text is the key to understanding consumer's attitude and experience of the products, and these raw textual contents need to be further cleaned to efficiently perform further analyses. We use R to clean and process review texts using following steps. First, we construct a corpus from review texts as the foundation for subsequent steps. Second, we capture relevant terms and remove irreverent and infrequent terms. To do so, we make all terms lowercase and remove numbers and special characters. Words with the same root form are combined into one term, because these terms have similar meanings. For example, "brewed", "brewer", "brewing", and "brew" are combined into "brew". We use *stop words* list to remove irrelevant terms. Specifically, we remove a list of basic and common terms (e.g., the, have, will, just) that are not useful for further analysis. To remove terms rarely used in reviews, we use the *RemoveSparseTerms* function at 0.99 level, which removes terms that occur less than 1% of all documents. This common practice allows us to obtain a shorter list of terms with more useful information, which results in 202 distinct words for the topic mining analysis[1]. The word frequency is presented in figure 1 using *Word Clouds*. The most frequently occurred words include “taste”, “flavor”, “great”, “love”, and “price”, describing different aspects of a review.

Methods

Latent Dirichlet Allocation

In this study, a widely-used topic model method, Latent Dirichlet Allocation (LDA) is applied to discover the underlining topics from the review text. LDA is a Bayesian generative model, it associates each of *D* document with a distribution over *K* latent topics, and each topic is a multinomial distribution over a *W* word vocabulary. Following Blei et al (2003), the graphical LDA procedure can be illustrated in figure 2. Define θ_i is the topic distribution for document *i*, φ_k is the word distribution for topic *k*, z_{ij} is the topic for the *j*th word in document *i*, and w_{ij} is a specific word. Denote *M* as the number of documents, *N* as the number of words in a document, α is the Dirichlet parameter on topic distribution over the words, and β are the Dirichlet parameter on the word distribution. For each document *d*, LDA goes through each word *w* in *d* and for each topic *k* and assumes the following generative process: 1) choose $N(words) \sim Poisson(\xi)$, 2) choose $\theta(topic\ distribution) \sim Dirichelet(\alpha)$, 3) for each of the *N* words w_n: a) choose a topic $z_n \sim Multinomial(\theta)$, b) choose a word w_n from $p(w_n|z_n\beta)$, a multinomial conditional on the topic z_n. LDA model can capture key information and important statistical relationship while reduce the complexity of the text corpus. The goal of the generative model is to find the group of topics that best describe the observed words in all the documents.

[1] Different levels of Sparse was specified, which results in similar topics when conducting the data analysis.

The information generated from the LDA model include the key words associated with each of the topics and the probability that each of the text review belong to each topic.

Regression Analysis

We use regression analysis to examine the interpretation power of hidden topics. The dependent variable is the votes for review helpfulness (*Helpfulness*). As dependent variable is a count variable, both Poisson and negative binomial regressions can be used for count data. However, the Poisson regression requires the mean to be equal to the variance, while the mean of helpfulness votes is smaller than the variance (table1). Therefore, we choose the negative binomial regression model to overcome this over-dispersion problem. Following Greene (2011), let $\boldsymbol{x_i}$ to be a vector of independent variables and $\boldsymbol{\beta}$ is a vector of parameters, the negative binomial model can be written as:

$$\mathrm{P}(\mathrm{Y} = y_i|\boldsymbol{x_i}) = \frac{\Gamma(\theta+y_i)}{\Gamma(1+y_i)\Gamma(\theta)}\gamma_i^{y_i}(1-r_i)^{\theta} \tag{1}$$

where $r_i = \frac{\lambda_i}{\theta+\lambda_i}$, $\lambda_i = \exp(\boldsymbol{x_i'\beta})$.

Because we are interested in testing the impact of underlying topics on helpfulness votes, we construct three regressions with the same dependent variable but different independent variables. The first regression only includes numeric review variables such as overall rating and review length, and the second regression includes both numeric review variables and topic dummies that indicate the implicit topics with which the customer review is associated with. The third regression has a non-linear form besides including both numeric variables and topic variables. AIC and log-likelihood ratio test are used to determine the better model.

Results

One of the most important and critical components of the LDA is to determine the number of topics that underlie all the text input. In this study, we used two metrics introduced by Cao et al. (2009) and Deveaud et al. (2014) to determine the optimal number of topics for the customer review texts. At the optimal number of the topics, the method by Cao et al. (2009) finds the minimization, whereas the method by Deveaud et al. (2014) finds the maximization of their matrices, respectively. We test a range of topic numbers from 2 to 15 and both methods indicate that four topics is the optimal choice (figure 2). We then use the LDAvis package to visualize the topic mapping. The visualized distribution of the four topics generated by LDA model is displayed in figure 4, which shows that each of the four topics is in its own non-overlapping region. This indicates that there is no correlation between the four topics, implying that each topic is unique and informative. Figure 5 show the topics and the corresponding percentage that each topic makes up for all reviews. Nearly one-third of the customer reviews is about topic 1, nearly one-fourth of reviews belong to topic 3, and topic 2 and topic 4 account for about one-fifth of total reviews, respectively.

A good topic model not only depends on the models' performance in measurable statistical metrics, but also on the reasonability and the interpretability of each topic. As a result of LDA, each topic is made up of terms/words; and in our study, the top 20 terms in each topic are selected (table 2). Since LDA allows for a word to be part of multiple topics, some words appear in more than one topics. For example, the word "pack" is part of both topic 2 and topic 4, and word "brand" appears in both topic 1 and topic 2. Results in table 2 show that Topic 1 is more likely to consist of words describing shopping experience at Amazon.com, such as "buy", "order", "purchase", "subscribe", and "shipping". Thus, we refer this topic as "Amazon service". Topic 2 contains terms describing "physical" aspects of coffee products and related machine products, such as “Keurig”, “pack”, “plastic”, and “machine”, and we call this topic as “physical feature”. The third topic consists of many subjective and emotional expressions, such as “taste”, “love”, “strong”, “recommend”, and “delicious”, so this group is referred as “subjective expression”. The last topic covers words describing the flavors and verities of coffee products, including words “flavor”, “bold”, “dark”, and “roast”. As noted that “donut” and “shop” are in topic 4, we checked the original reviews mentioned these two words and found that most of them talked about one flavor called “donut shop”. Therefore, we call topic 4 as “flavor feature”. Therefore, all the four distinct topics generated from the review text are interpretable and reasonable.

Now we examine the impact of the four topics on helpfulness votes for customer reviews. The estimated results for the negative binomial model with three sets of variables are presented in table 3. Model 1 only consists of two numeric review variables, overall rating and the length of review that has been examined extensively by previous studies. Both variables are statically significant at 1% level. The coefficient on *Rating* is negative, indicating that reviews with lower overall ratings are more likely to obtain votes for helpfulness. The coefficient on *Length* is positive, indicating the longer the reviews are more likely to be perceived as helpful. Such results are consistent with previous findings. Mudambi and Schuff (2010) found that for experience goods, there is a significant negative relationship between rating and helpfulness. Hao et al. (2009) and Kim et al. (2006) found the review length has a positive effect on perceived helpfulness of consumer reviews.

Model 2 includes both numeric review variables and topic variables. Considered variables are all statistically significant at 5% level. To avoid singularity, topic 1 “Amazon service” is dropped from the estimation to serve as the base. The coefficients on *Rating* and

Length are comparable with those in Model 1. Topic 2 “physical feature” and topic 4 “flavor feature” have positive coefficients, while topic 3 “subjective expression” has a negative coefficient. Such results show that although all underlying topics have a significant impact on the helpfulness votes, they influent the dependent variable in two different directions. The baseline “Amazon service” is somewhat exogenous from coffee products themselves. Compared to Amazon service, reviews discussing coffee flavors, varieties, and related products (e.g., coffee machine) are perceived as more helpful by review readers. Meanwhile, reviews discussing personal emotions and feelings are less valued by review readers. Such results are supported by Connors et al. (2011), who pointed out that reviews appearing too emotional or biased and lack of objective information are perceived unhelpful.

Additional to the variables in Model 2, Model 3 include variable length squares to capture the nonlinear effect of review length. Results show a strong nonlinear relationship between the review length and the helpfulness vote (table 3). A longer review provide more information on the product, therefore, make the review more helpful. However, when review text become too long, it starts to carry some information that is not helpful or shoppers may be overwhelmed by the large amount of information provided in the review text, therefore trying to skip or ignore some review text, thinking it less helpful. Figure 6 shows that after the number of characters in a review increase to about 2,291 characters, more text starts to have a negative impact on the helpfulness of the review. This result is consistent with the general consensus regarding the impact of information that too much information may confuse consumers and result in less optimal choice (Jacoby et al. 1974; Malhotra 1984). Results in Model 3 also shows that the “physical feature” of the coffee become insignificant, equally important as the “Amazon service” in providing helpful information, while the “subjective expression” and “flavor feature” still have significant negative and positive impact, respectively on the helpfulness of the review text.

As shown in table 3, Model 2 has a smaller AIC value than Model 1, while Model 3 has a smaller AIC value than Model 2. The log-likelihood ratio test indicates that Model 2 is preferred to model 1 and Model 3 is preferred to Model 2. These results indicate that topic variables can improve the interpretation ability for the helpfulness votes and lead to a better performance of the model. Moreover, the model can be significantly improved if the nonlinear effect of the review *Length* is considered. As for the relative importance of the explanatory variables, because the variables in the model are not measured on the same scale, their impact on the helpfulness votes is not directly comparable. However, the results demonstrate that if the overall *Rating* of a product increases by one point, the rate for helpfulness votes would be expected to decrease by a factor of -0.363. Increasing length of the review will increase the helpfulness rating first, but after the length increase to more than 2,291 characters, more text starts to hurt the helpfulness of the review. Compared to the reviews that focus more on the “Amazon service”, reviews that centralized on “physical feature” provide no more helpful information. At last, reviews characterized by “subjective expression” decrease the helpfulness of the reviews significantly while reviews focusing on “flavor feature” significantly increase the helpfulness of the reviews and at a large scales (0.241 vs. -0.138).

Conclusion

Online shopping has become a major channel for consumers to purchase a wide range of products worldwide, and food and grocery products have also become a part of consumers’ online shopping choices. Given this trend, consumer reviews do not only provide a platform for consumers to share their opinions and experiences but for potential consumers to obtain knowledge about the product to make purchase decisions. Moreover, textual reviews themselves provide e-commerce companies a collection of information about consumers' preferences and opinions of the products without introducing any biases via researcher or company designed surveys. Customer reviews have been investigated for various products, including digital cameras, cell phones, software programs, and hotels. However, little is known about food products. The food product is a unique experience good, and perceived quality can be highly subjective and heterogeneous. Therefore, how review readers obtain desirable information on food products becomes an interesting question. In addition, many previous studies focused on numeric review variables rather than the contents provided by review text, which could lead to a less efficient evaluation of the review system. Our study aims to fill the gap in the literature and explores the consumer perceptions and preferences from online food reviews.

The helpfulness vote feature is a tool for consumers to identify useful information and reduce search costs when available information is overwhelming. When we researchers receive a large amount unstructured information, we need to rely on advanced analytical techniques to discover hidden information. In this study, we use two major machine learning tools, text mining and topic mining modeling such as LDA to explore the textural content and discover hidden topics behind the reviews.

Our analysis reveals four interpretable topics underlying all the text reviews of coffee: "Amazon service", "physical feature", "subjective expression", and "flavor feature". Including the topics as explanatory variables significantly improves the power of regression models that explain the variation in the helpfulness votes. However, the different topics influence helpfulness votes in quite different directions. Specifically, coffee review readers perceive information on objective aspects of coffee (e.g., flavor feature) as more helpful; the reviews that objectively discussing related products (e.g., coffee brewing machine) equivalently valuable as those focusing on Amazon services and shopping experiences. On the other hand, reviews that talk more about personal emotions and feelings are perceived as less helpful. This might because food evaluation is usually subjective and depends on personal preference. Therefore, one person's taste and emotion associated with a product do not provide much useful information for another person.

With the development of technology, it is expected that the sale of online grocery shopping would keep increasing. The huge potential in online grocery shopping not only attracts traditional retail giant like Walmart (Walmart Grocery) and online retailer leader such as Amazon (AmazonFresh), but also intrigue upstarts, such as Instacart and FreshDirect. Despite the fasting growth in this area, much less is known regarding the important factors that may affect consumer online grocery shopping behavior compared to other popular products such as electronics, books, etc. There is a critical need for research in this area because groceries such as food not only provides services that meet a personal need, it also affect peoples' health, therefore bearing a social impact. By using machine learning techniques and Amazon review data, our discovered the important factors that influence the helpfulness of information provided in Amazon text review. Our results imply that retailers should encourage their customers to provide more unbiased reviews that focus on the products itself, and discourage the reviews that are too personal and emotional. From a broader policy perspective, our results demonstrate the importance of the diffusion of relevant information, such as "objective feature" in assisting consumers to make a decision. Development of platforms that encourage the provision of the objective opinions and discussion of products may help consumers make better food choices. Because the underlying topics could be subject to studied products, future studies can test different food categories to discover the generated topics among all food products. Also, we do not have information on review authors. If this information is available, studies can associate topics with consumer segmentations, which would provide detail information to help develop effective strategies to communicate with target consumers.

Reference

Archark, N. A. Ghose, and P.G. Ipeirotis. 2011. “Deriving the pricing power of product features by mining consume reviews.” *Management Science*, 57(8): 1485-1590.

Berger, J., A.T. Sorensen, S.J. Rasmussen. 2010. “Positive effects of negative publicity: when negative reviews increase sales.” *Marketing Science*, 29(5): 815-827.

Blei, D., A. Ng, M. Jordan. 2003.“Latent Dicichlet Allocation.” *Journal of Machine Learning Research*, 2:9931022.

Cao, Q., W. Duan, and Q. Can. 2011. “Exploring determinants of voting for the ‘helpfulness’ of online user reviews: A text mining approach.” *Decision Support Systems* 50: p.511-521.

Cao. J, T. Xia, J. Li, Y. Zhang, and T. Sheng. 2009. “A density-based method for adaptive LDA model selection.” Neurocomputing-16th European Symposium on Artificial Neural Networks, 72:1775-1781.

Chevalier, J.A. and D. Mayzlin. 2006. “The effect of word of mouth on sales: online book reviews.” *Journal of Marketing Research*, 43: p.345-354.

Chen, P., S. Dhanasobhon, and M. Smith. 2008. “All reviews are not created equal: the disaggregate impact of reviews on sales on Amazon.com.” available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=918083.

Connors, L., S.M. Mudambi, and D. Schuff. 2011. “Is it the review or the reviewer? A multi-method approach to determine the antecedents of online review helpfulness.” Hawaii International Conference on System Sciences, IEEE, p1-10.

Deveaud, R. E. SanJuan, and P. Bellot. 2014. “Accurate and effective latent concept modeling for ad hoc information retrieval.” *Document numerique*, 17(1): 61-84.

Duan , W., B. Gu, A.B. Whinston. 2008. “Do online reviews matter? – An empirical investigation of panel data.” *Decision Support System*, 45(10): p.1407-1424.

Felbermayr, A. and A. Nanopoulos. 2016. “The role of emotions for the perceived usefulness in online customer reviews.” *Journal of Interactive Marketing*, 36: p.60-76.

Forrester. 2016. “US cross-channel retail forecast, 2015-2020.” Available at: https://www.forrester.com/report/Brief+US+CrossChannel+Retail+Forecast+2015+To+2020/-/E-RES116715 (accessed on January 10, 2017).

Greene, W.H. 2011. Econometrics analysis (7th ed). Upper Saddle River, NJ: Prentice Hall.

Hao, Y., Y. Li, and P. Zou. 2009. “Why some online product reviews have no usefulness rating.” Pacific Asia conference on information systems, Hyderabad, India, p.100.

Huang, J., S. Rogers, and E. Joo. 2013. “Improving restaurants by extracting subtopics from Yelp reviews.” Available at:

https://www.yelp.com/html/pdf/YelpDatasetChallengeWinner_ImprovingRestaurants.pdf (accessed on January 6, 2017).

Jacoby, J., Speller, D.E. and Kohn, C.A., 1974. Brand choice behavior as a function of information load. *Journal of Marketing Research*, pp.63-69.

Kim, S.M., P. Pantel, T. Chklovski, M. Pennacchiotti. 2006. "Automatically assessing review helpfulness." Conference on Empirical Methods in Natural Language Processing, Morristown, NJ, p.423-430.

Lee, T.Y. and E.T. Bradlow. 2010. "Automated marketing research using online customer reviews." *Journal of Marketing Research*, 48(5): 881-894.

Malhotra, N.K., 1984. Reflections on the information overload paradigm in consumer decision making. *Journal of Consumer Research*, 10(4), pp.436-440.

McAuley, J. and J. Leskovec. 2013. "Hidden factors and hidden topics: understanding rating dimensions with review text." Proceedings of the 7th ACM conference on Recommender systems, ACM: p.165-172.

McAuley, J., C. Targett, Q. Shi, and A.V.D. Hengel. 2015. "Image-based recommendations on styles and substitutes." Proceedings of the 38th International ACM SIGIR conference on Reseach and Development in Information Retrieval. ACM, p43-52.

Moon, S. and W.A. Kamakura. 2017. "A picture is worth a thousand words: Translating product reviews into a product positioning map." *International Journal of Research in Marketing*, 34: p.265-285.

Mudambi, S.M. and D. Schuff. 2010. "What makes a helpful online review? A study of customer reviews on Amazon.com." *MIS Quarterly*, 34: p.185-200.

Netzer, O., R. Feldman, J.Godenberg, and M. Fresko. 2012. "Mine your own business: Market structure surveillance through text mining." *Marketing Science*, 31(3): 521-543.

Nielsen. 2015. "E-commerce, digital technology and changing shopping preferences around the world." Available at: https://www.nielsen.com/content/dam/nielsenglobal/vn/docs/Reports/2015/Nielsen%20Global%20E-Commerce%20and%20The%20New%20Retail%20Report%20APRIL%202015%20(Digital).pdf (accessed on January 10, 2017).

Nelson, P. 1970. "Information and consumer behavior." *Journal of Political Economy*, 78(20): p.311-329.

US Census. 2017. "Quarterly retail e-commerce sales 1st quarter 2017." Available at: https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf (accessed on May 16, 2017).

Table 1. Descriptive statistics for numeric review variables

Variable	Mean	Std.Dev	Min	Max
Rating	4.40	1.08	1	5
Length	258.33	251.13	4	5003
Helpfulness	0.59	7.93	0	590

Table 2. Top 20 terms in each topic

Topic 1	Topic 2	Topic 3	Topic 4
Price	Cup	Taste	Flavor
Buy	Keurig	Great	Roast
Order	Brew	Love	Blend
Good	Box	Strong	Bold
Pod	Pack	Good	Favorite
Product	Plastic	Drink	Variety
Amazon	Work	Bitter	Donut
Purchase	Review	Morning	Dark
Brand	Bag	Recommend	Shop
Find	Fresh	Smooth	Pack
Starbucks	Machine	Perfect	Enjoy
Store	Water	Rich	Medium
Senseo	Regular	Husband	Decarf
Maker	Open	Delicious	Differentials
Subscribe	Brand	Nice	Breakfast
Excellent	Sanfranciscobay	Smell	French
Shipping	Small	Aroma	Weak
Month	Fog	Start	People
Delivery	Grands	Wonderful	Greenmontain
Quality	Bad	House	Brooklyn

Table 3. Parameter estimates for negative binomial models

	Model 1		Model 2		Model 3	
Variable	Coefficient	Std.error	Coefficient	Std.error	Coefficient	Std.error
Intercept	-0.550***	0.094	-0.559***	0.100	-0.817	0.105
Rating	-0.365***	0.020	-0.366***	0.020	-0.363	0.020
Length	4.000***	0.000	0.003***	0.000	5.000***	0.000
Length^2					-0.001***	0.000
Topic 2			0.136**	0.065	0.088	0.065
Topic 3			-0.159**	0.066	-0.138**	0.066
Topic 4			0.263***	0.065	0.241***	0.065
AIC	28179		28147		28050	
Log-likelihood	-14085.51		-14066.6		-14017.1465	
Likelihood Ration Test	Model 1 vs Model 2 $\chi^2 = 37.8$, Pvalue=0.000			Model 2 vs Model 3 $\chi^2 = 98.95$, Pvalue=0.000		

Note: *, **, and *** indicate significance at 10%, 5%, and 1% significance levels, respectively.

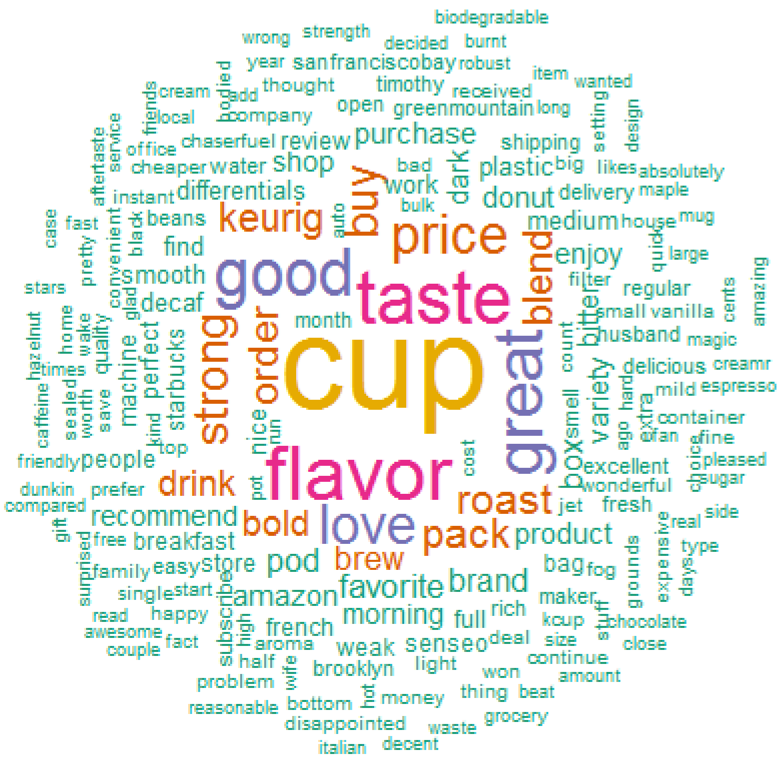

Figure 1. Word clouds of word counts in coffee reviews

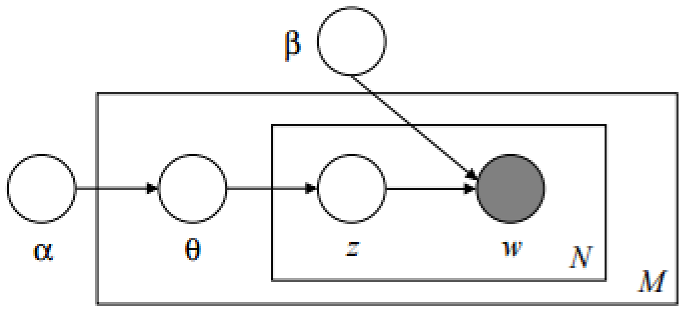

Figure 2. Graphical model representation of LDA (Source: Blei et al. [2003])

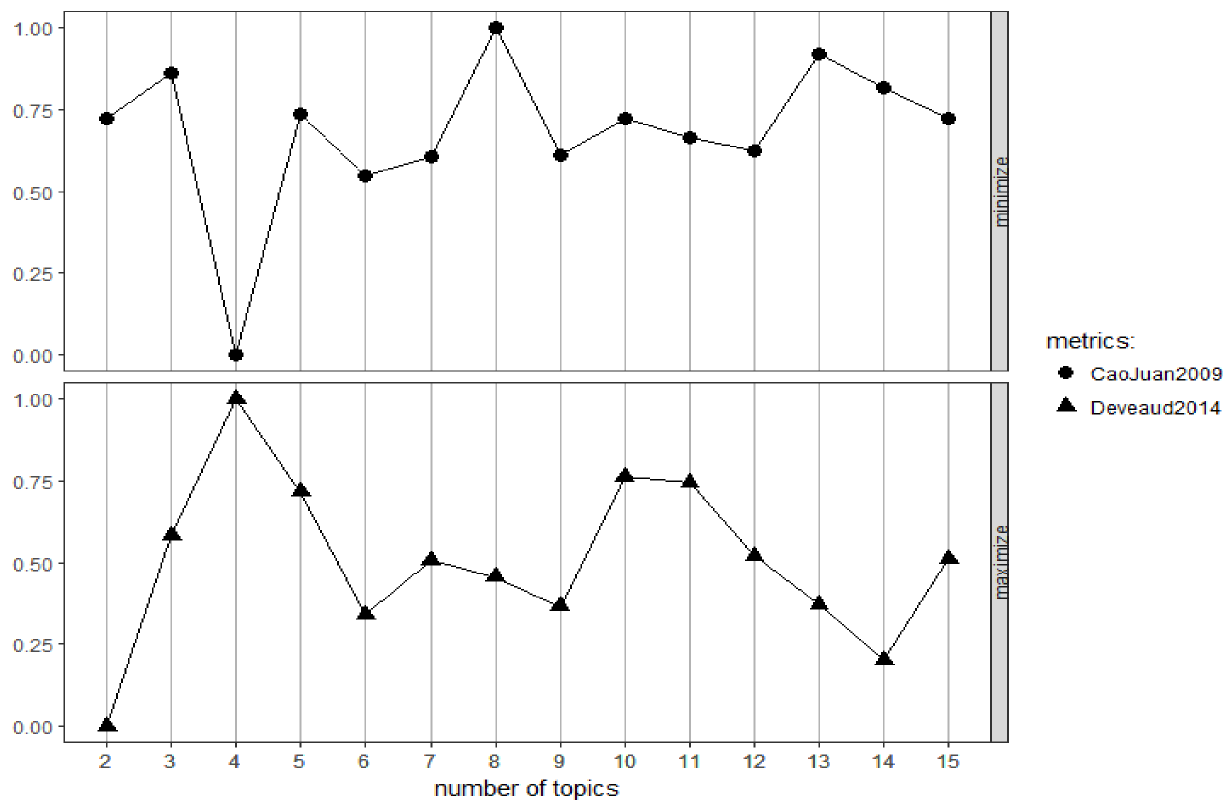

Figure 3. Selection of the number of topics

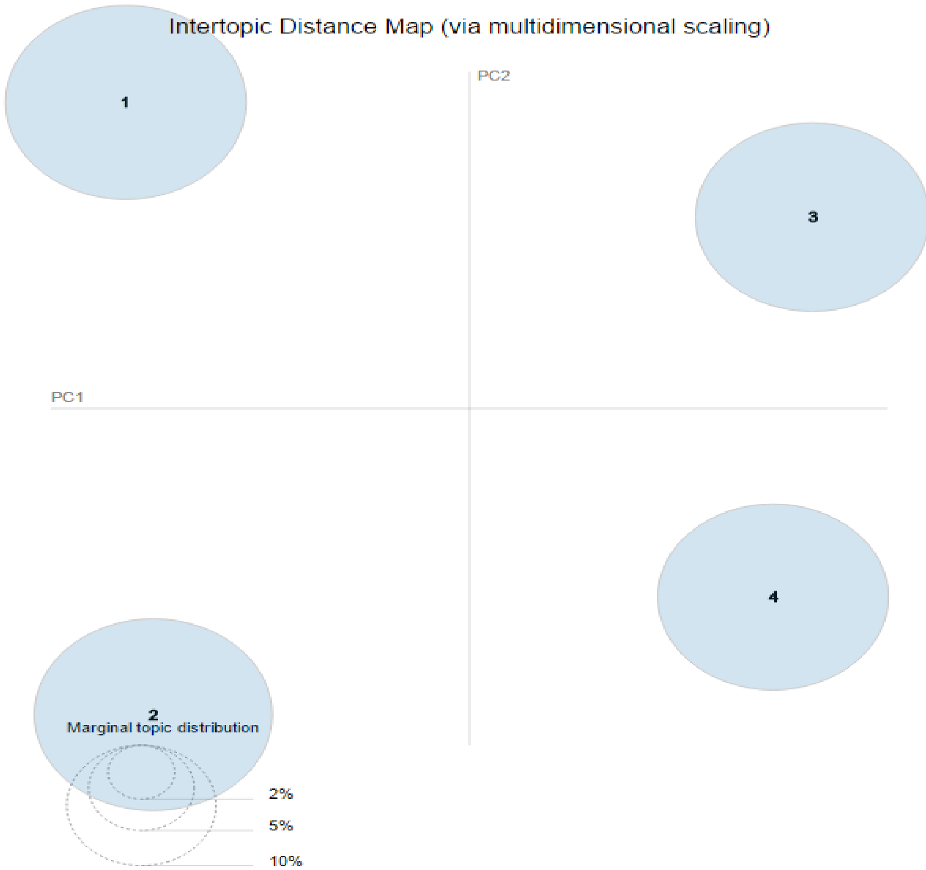

Figure 4. Visualization for topic distribution

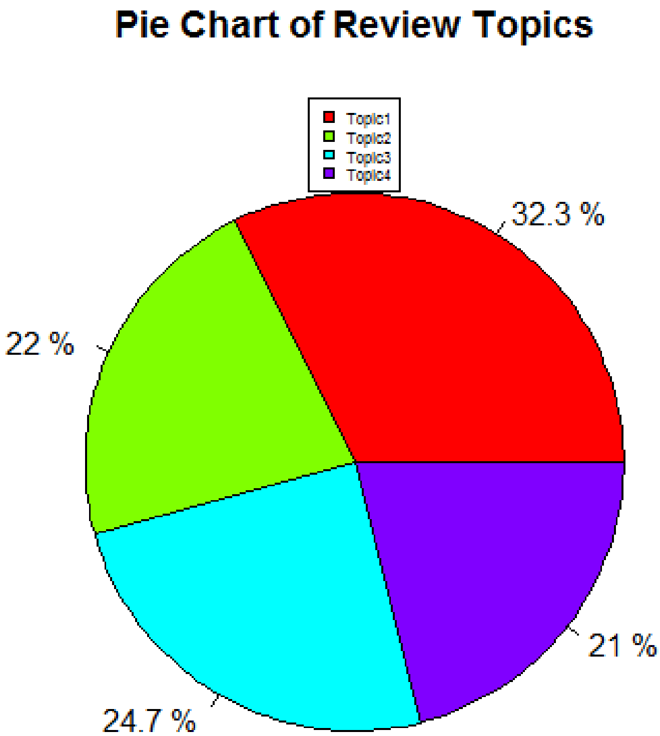

Figure 5. Pie chart for topics

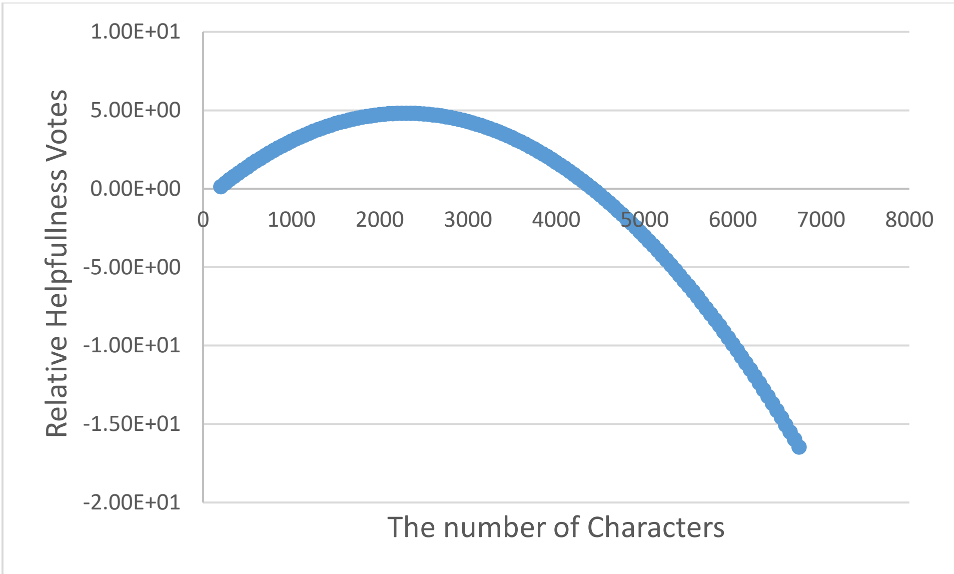

Figure 6. Changes in Helpfulness Votes with the Number of Characters in Review Text