



## Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach

Yan Heng, Zhifeng Gao\*, Yuan Jiang, Xuqi Chen

Food and Resource Economics Department, University of Florida, USA



### ARTICLE INFO

#### Keywords:

Online coffee shopping  
Customer reviews  
Topic modeling

### ABSTRACT

Despite the popularity of online food and grocery shopping, little research has been conducted to understand the factors that influence consumers' online food purchases. Using a topic modeling approach, our results show four interpretable factors have significant impacts on the helpfulness of customer reviews: Amazon Service, Physical Feature, Flavor Feature, and Subjective Expression. Readers of customer reviews perceive objective reviews as more helpful than subjective reviews. In addition, customer review helpfulness has a concave relationship with the length of the reviews. Our results provide important business implications on how to encourage more helpful reviews to assist potential shoppers in making better purchase decisions.

### 1. Introduction

Understanding consumer preferences of food products is vital for food industries and retailers to develop optimal production and marketing strategies. With the increasing popularity of online food and grocery shopping, there is a critical need for better information on the factors that affect consumers' choices of online food products. Different from surveys commonly used to obtain information on consumers' preferences and experiences, online customer reviews provide a rich source of information without introducing any biases by researchers or company surveys. Using the plethora of consumers' review information provided by Amazon's online customers of the products they purchase (customer reviews), we applied a topic modeling approach to determine the hidden factors that may affect consumers' purchase decisions of different brands of coffee. Our approach shows that the helpfulness count of customer reviews is highly associated with the implicit factors that affect consumer purchase decisions during online shopping. The approach also has a wide application to research focusing on the factors affecting consumer online food and grocery shopping behavior of other products.

#### 1.1. Background

Online stores have become an essential channel for sales of a wide range of products. Over 60% of respondents in the Asia, Africa/Middle East, and Latin America regions are willing to shop online in the future (Nielsen.com, 2015). In the first quarter of 2017, total U.S. retail e-commerce sales reached US\$105.7 billion, an increase of 4.1% from the

fourth quarter of 2016 (U.S. Census Bureau, 2017). Online sales are expected to grow at an annual rate of 9.3% by 2020 (Forrester.com, 2016). While online grocery shopping only accounts for a small proportion of the market, online food and grocery shopping is growing globally. According to a Nielsen report, 25% of respondents surveyed have purchased groceries online, and over fifty percent are willing to do so in the future (Nielsen.com, 2015). Online grocery sales are projected to reach nearly US\$100 billion by 2019 in the United States (Kumar, 2014). Mintel.com (2016) reported that Millennials in Europe have led the online grocery shopping revolution and nearly 50% of younger consumers (aged 16–24) have shopped online for groceries. The recent acquisition of Whole Foods Market by Amazon is a good example reflecting this trend, showing the future potential of online food and grocery shopping. Because consumers' purchase decisions affect sales directly, understanding how consumers make decisions during online shopping has become an essential subject for both researchers and e-commerce companies.

Studies that explored the factors influencing consumers' online shopping decisions found that online consumers' preferences vary by product category (Nilsson et al., 2015; Schultz and Block, 2015). Consumers' demographics influence shopping behavior and preferences, with convenience being more appealing to men than to women, especially in the 34–50 age group (Passyn et al., 2011). Liu et al. (2015) clustered online consumers into economical purchasers, active-star purchasers, direct purchasers, high-loyalty purchasers, risk-averse purchasers, and credibility-first purchasers. They concluded that economical purchasers are most sensitive to discount promotions; active-star purchasers are most sensitive to word-of-mouth promotions; and

\* Correspondence to: University of Florida, 1155 MCCA Hall, Gainesville, FL 32611, USA.  
E-mail address: [zfgao@ufl.edu](mailto:zfgao@ufl.edu) (Z. Gao).

direct purchasers are most sensitive to advertising promotions. Clemes et al. (2014) conducted a study in China and ranked the important decision factors affecting online shopping as perceived risk, consumer resources, service quality, subjective norms, product variety, convenience, and website factors. A large body of literature demonstrated that customer reviews are one of the most influential factors affecting consumer online shopping decisions (Chen et al., 2008; Duan et al., 2008; Engler et al., 2015).

### 1.2. The role of customer reviews

The overwhelming number of competing products and overloaded information may make it difficult for online consumers to make choices. Online retailers' websites, such as Amazon.com, provide consumers a platform to post reviews to share their opinions and experiences with the products they have purchased. These reviews, in turn, attract more customers and help potential consumers make better decisions (Cao et al., 2011). Although consumers generally would benefit from more information, many products have hundreds and even thousands of reviews with various content and inconsistent opinions, which makes it difficult for consumers to use the information effectively for decision making. In this sense, online retailers often highlight valuable information by creating the "helpfulness" feature, which allows consumers to evaluate other users' reviews. For example, Amazon.com posts the question, "Was this review helpful to you?" under each online review and tallies the votes for the convenience of its shoppers. Websites also prioritize online customer reviews based on helpfulness votes to reduce consumers' time in finding the useful information. Helpfulness votes, along with consumers' reviews, provide a large amount of information that enables researchers to examine factors that may affect consumers' purchase decisions in an online shopping environment.

Many studies have examined the factors that influence review helpfulness. For example, review rating (Mudambi and Schuff, 2010; Cao et al., 2011), the length of words (Kim et al., 2006; Hao et al., 2009), positive and negative sentiment (Berger et al., 2010), and emotions (Felbermayr and Nanopoulos, 2016) have been used to explain and predict the helpfulness votes. While the main focus of these studies was improving the review system design to encourage more helpfulness votes, there are no consistent conclusions regarding the important factors affecting the helpfulness of online reviews.

### 1.3. The motivation of this study

In this study, we hypothesize that helpfulness votes reflect consumers' demand for information that matters to them but is hidden behind the review text. Textual information contained in reviews is more important to consumers than review length or a product's overall rating (Chevalier and Mayzlin, 2006). Ignoring the text content of reviews is a major shortcoming of existing studies on recommender/recommendation systems (McAuley and Leskovec, 2013). In other words, all customer reviews are text information with high dimensionality and multiple latent interpretable topics underlying the texts determine the helpfulness of reviews. For example, Huang et al. (2013) show that Yelp reviews on restaurants can be categorized into multiple latent topics such as service, value, décor, and healthiness. In this regard, the high-dimensional review can be processed by topic mining approaches to extract the low-dimensional latent topics.

Further differentiating from previous studies, we examine the factors affecting the helpfulness of customer reviews of food products that are rarely studied in this area. Products can generally be categorized into search goods and experience goods. A search good is a product whose quality is observable before purchase; an experience good is a product whose quality can only be revealed after consuming or experiencing it (Nelson, 1970). Previous studies have examined factors affecting the helpfulness of customer reviews on both search (e.g. digital cameras) and experience goods (e.g. software programs, sedan

cars) (Archark et al., 2011; Cao et al., 2011; Lee and Bradlow, 2011; Netzer et al., 2012). For example, in comparing customer reviews on search goods (digital cameras, cell phones, and laser printers) and experience goods (music CDs, MP3 players, and video games), Mudambi and Schuff (2010) found that product type has a significant impact on the helpfulness of customer reviews.

While the products mentioned above have well-defined standards for product attributes and evaluations, food is a unique category of experience goods. Consumers' preferences for food are heterogeneous, so it is not easy to consistently define the important attributes associated with food products. Moreover, the perceived quality of food can be highly subjective, and this is particularly true for products heavily characterized by their sensory attributes (Moon and Kamakura, 2017). In addition, food directly relates to human health and welfare. With customer review systems becoming communication tools to affect consumers' perceptions and behavior in online shopping, exploring online customer reviews would be helpful to identify potential food risks and benefits as well as to convey this information to the public. Until recently, few studies have focused on customer reviews of food products except for Moon and Kamakura (2017) who showed that wine reviews were more likely to rely on experts' opinions. This makes wine different from most food products that rely heavily on regular consumers' evaluation.

To fill the gap in the literature, our study aims to discover multiple latent interpretable topics in consumers' online text reviews of coffee, and further explore how these topics may influence online shoppers' decision.

## 2. Data

We use Amazon's customer review data on grocery products from 2004 to 2014 collected by McAuley et al. (2015). In the dataset, there are a total of 1,048,576 distinct reviews, which include reviewer information on reviewer's ID, product ASIN code, review helpfulness vote, review text, product overall rating, and review date. We selected coffee as the target product in this study because coffee has the largest number of reviews among all the food products, and coffee is a good example of sensorial food products with various differentiations. The data used for analysis include 22,424 customer reviews of coffee.

For each customer review, we generate the following variables: the total number of people who voted a review as helpful (*Helpfulness*), the overall rating of the product (1–5) given by the reviewer (*Rating*), the word count of the review (*Length*), and the difference in time between 2015 and the year when the review was posted (*Time*). We use the helpfulness votes as the dependent variable in the regression models in the next section. The overall rating and review length are used as independent variables based on the literature (Kim et al., 2006; Hao et al., 2009; Mudambi and Schuff, 2010). The overall rating provides readers a quick impression of a consumer's evaluation of the product. The review length indicates the amount of the information supplied by the review text. Time is considered because a longer elapsed time may increase the helpfulness votes that a review could receive (Fang et al., 2016). The descriptive statistics are presented in Table 1. The overall rating of coffee products is positive, with an average rating of 4.40. On average, each review contains about 48 words and scores a 0.59 helpfulness vote. The number of helpfulness votes varies from zero to 590

**Table 1**  
Descriptive statistics for numeric review variables.

| Variable           | Mean  | Std. Dev | Min | Median | Max |
|--------------------|-------|----------|-----|--------|-----|
| <i>Rating</i>      | 4.40  | 1.08     | 1   | 5      | 5   |
| <i>Length</i>      | 48.26 | 45.89    | 1   | 32     | 872 |
| <i>Helpfulness</i> | 0.59  | 7.93     | 0   | 0      | 590 |
| <i>Time</i>        | 4.56  | 1.62     | 1   | 4      | 11  |

Download English Version:

<https://daneshyari.com/en/article/7433553>

Download Persian Version:

<https://daneshyari.com/article/7433553>

[Daneshyari.com](https://daneshyari.com)