Modeling Shopping Cart Decisions with Heterogeneous Product Involvement and Reviews

Research-in-Progress

Hongming Gao **Carol X.J. Ou** **Hongwei Liu**

Abstract

The most recent consumer propensity study by SAP indicates that online shopping cart abandonment is high and the associated reasons are complex. In order to examine this phenomenon, we construct online SCA decision as a discrete choice model (DCM) and capture consumer segments by a latent class model (LCM) in this research-in-progress (RIP) paper, grounded on the theories of product involvement, word of mouth, and consumer heterogeneity. We will apply the clickstream dataset from 78,746 consumers at a large Chinese online platform to verify the proposed models in future study. The objective of this research project is to scrutinize the heterogeneous impacts of product involvement and online reviews on shopping cart decision-making in view of individual-level sequential behavior and the associated products in the form of stock-keeping-unit items. We conclude this RIP paper with the discussion of potential theoretical contributions and managerial implications.

Keywords: Shopping cart decision-making, shopping cart abandonment (SCA), consumer involvement, clickstream data, consumer heterogeneity

Introduction

Despite the surge of online sales, shopping cart abandonment (SCA) has become a growing concern for Internet retailers. The most recent SAP Consumer Propensity study[1] of 600-1000 consumers in each of the 24 global markets reveals that online SCA was high, hovering between 41% to 57% in 2018. These numbers suggest how to effectively and strategically decrease the online SCA rate already became one of the utmost interests of e-commerce retailers. As remarked by Jennifer Arnold, the vice president of marketing at SAP, "reviewing cart abandonment data provides a starting point for retailers to identify friction points in the consumer journey and make improvements to the overall purchasing experience"[2].

From the perspective of the existing research on online consumer decision-making process, purchasing behavior and non-purchasing behavior can be considered separately. However, the majority of extant literature focuses on purchasing behavior, e.g., product or brand sales performance, whether consumers buy or not, the number of repeated purchases (e.g., Gu et al. 2012; Kim et al. 2004; Mudambi and Schuff 2010; Ou and Chan 2014), with only few exceptions on investigating the influencing factors of SCA (e.g., Berry et al. 2002; Ding et al. 2015; Li and Chatterjee 2006; Kukar-Kinney and Close 2010). These studies have provided a good starting to understand consumer decision-making behavior of SCA, outlining that different effects of hedonic- or utilitarian- orientation of using shopping cart, and of stimuli and information content while browsing web pages. However, it is challenging to securitize the concrete consumer decision-making process due to the lack of research at the level of stock-keeping-

[1] Retrieved from www.thepaypers.com/ecommerce/shopping-cart-abandonment-significantly-high-in-australia/774287-25
[2] Retrieved from www.bandt.com.au/marketing/study-57-aussies-abandon-virtual-shopping-carts

unit (SKU) products (e.g., Ding et al. 2015; Li and Chatterjee 2006; Schellong et al. 2016) and individual behavior data (e.g., Berry et al. 2002;Kukar-Kinney and Close 2010).

Product involvement is determined by the level of a consumer's enduring perceptions and the extent of information needed directly toward a product to a shopping or hedonic end (Bian and Moutinho 2011; Richins and Bloch 1991; Schellong et al. 2016; Zaichkowsky 1986). Existing studies generally focus on traditional marketing factors on consumer decision-making process, such as online advertising and price promotion on web pages instead of considering specific product (Ding et al. 2015; Li and Chatterjee 2006; Schellong et al. 2016). The commonly ignored human-computer-interaction (HCI) factors such as clicking, browsing time, tagging behavior can implicitly both reflect the level of product involvement (Jiang et al. 2010; Schellong et al. 2016; Zaichkowsky 1986). When investigating consumers' clickstream data, product involvement can be measured by three quantitate variables, including cumulative browsing time duration, cumulative number of clicks, and if the product is tagged at the SKU level. Consumer responses have been different because of consumer heterogeneity. ***Consumer heterogeneity*** refers to consumers' diverse preference to differentiate consumer segments, market niches and product offerings (Allenby and Rossi 1998). We quantify the latent consumer heterogeneity to achieve consumer segments, since estimates of involvement effect might be biased if heterogeneity is not accounting for (Chatterjee et al. 2003; Ding et al. 2015; Li and Chatterjee 2006).

Online reviews refer to consumers' online comments based on their purchasing or/and product usage experience, which serve as quality signals in consumer purchase decisions (Gu et al. 2012). For potential consumers, such information reveals the quality of a specific product to influence decision-making. High- or low- involvement consumers are affected depending on their information processing ability (Lee et al. 2008). Nonetheless, category of high- or low involvement product varies from person to person instead of being subjectively defined (e.g. Gu et al. 2012; Park et al. 2007). Thus, we aim to investigate the effect of online review of a specific product on consumer involvement and shopping cart decision. Consistent with previous studies (Gu et al. 2012; Lee et al. 2008; Park et al. 2007), we focus on the number of online reviews and proportion of negative reviews.

This paper attempts to integrate the two separate research streams on consumers' purchase and non-purchase behaviors in order to unpack the black box of consumer decision-making process. In particular, we attempt to examine (1) *how consumers' shopping cart decisions are affected by the level of product involvement with the consideration of consumer heterogeneity?* (2) *how online product reviews might help consumers make shopping cart decisions?* Overall, we expect the holistic approach of simultaneously examining consumers' purchase and non-purchase behaviors can offer a better information management of consumer clickstream data from both theoretical and practical perspectives.

This RIP paper is structured as follows. After this introduction, the next section reviews the past studies on consumer decision-making process associated with product involvement, consumer heterogeneity and online reviews. The third section models the virtual shopping cart decision and consumer heterogeneity. We then explain the data set and the associated variables in the proposed models. We conclude the RIP paper with the discussion of expected contributions and implications.

Literature Review

Compared to traditional offline retailing, online retailers attract consumers to make purchase decisions mostly depending on information over the product itself (Detlor et al. 2003). That is, consumers search, evaluate and process online product/brand information in the pre-purchase stage and such information subsequently helps consumers make shopping cart decisions as illustrated in the Figure 1. For instance, consumer A first clicks (action acronym as C in Figure 1) and views the product 2 in a short time. Next she tags product 3 into her favorite sets (action acronym as F) and browses the webpage of product 3 for a relatively long time, finally discards product 3 from shopping carts (action acronym as D) as shown in Figure 1a.

Clickstream data as illustrated by Figure 1 can be recorded by web services or third-party services (Bucklin et al. 2009). This kind of data analysis method uses the objective consumer log data from webs browsing behavior. Involvement is the construct describing the perceived relevance of a person toward to the object based upon inherent cognitive demands, values and interests (Zaichkowsky 1986). Further,

product involvement refers to the level to which individual consumer is consciously involved or engaged with the product website and facilitates acquiring product information to form consideration set which meets his/her inner demand before making a virtual online shopping cart decision (Chatterjee et al. 2003; Schellong et al. 2016). The above HCI behavior as shown in Figure 1a can implicitly indicate a consumer's preference of a product or brand (Bucklin and Sismeiro 2009; Chatterjee et al. 2003), as well as reflect the level of product involvement in a consumer's psychological account (Jiang et al. 2010; Schellong et al. 2016). Highly involved products mean that consumers are more willing to be engaged with products to make the right decisions.

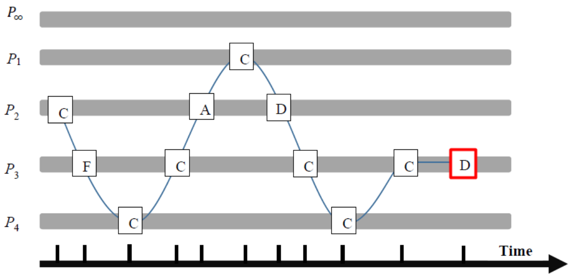

Fig 1a. Consumer A discards product P_3

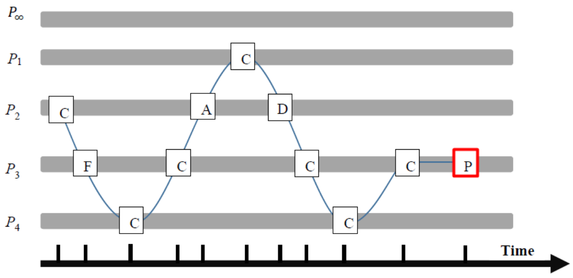

Fig 1b. Consumer B purchases product P_3

Figure 1. Similar Online Browsing Paths but Different Final Decisions

Notes: Symbols C stands for ***Click***, F for tag to ***Favorite*** sets, A for ***Add*** to shopping cart, D for ***Discard*** shopping cart, P for ***Purchase***. The grey bars indicates various products and the number of items is assumed to be sufficient. The horizontal axis means product combination set. Lateral axis represents time period, and a grey bar displays occurring time of a corresponding interactive behavior. The interval of two grey bars indicates the period of browsing time. For instance, consumer A in Figure 1a clicks product 2 first and then tags product3 into favorite sets, next clicks product 4. As shown in Figure 1, the time spending on product 3 is relatively longer than the time interval for product 2.

We argue that the examination of the flow of such clickstream data can help understand and identify consumers' purchasing intentions. Practically e-commerce platforms can provide more timely and tailor-made product recommendations along the way of consumer browsing. In reality, it is possible to have another customer, namely consumer B in Figure 1b, has extremely similar browsing behavior but at the end of the shopping journey decides to buy product 3. Consumer responses can be different, which might be partially explained by consumer heterogeneity (Allenby and Rossi 1998; Chatterjee et al. 2003; Ding et al. 2015; Li and Chatterjee 2006). Heterogeneity is a nuisance problem in the application of econometrics, has to be coped with (Allenby and Rossi 1998). Once the heterogeneity is neglected that parameter estimation of product involvement might be biased. Prior literature displays that capturing consumer heterogeneity would be more accurate to interpret the nature of consumer decision-making (Chatterjee et al. 2003; Ding et al. 2015; Li and Chatterjee 2006).

In addition to the information about product descriptions, online review is considered to be able to make or break a consumer's purchasing decisions (Huang et al. 2018; Mudambi and Schuff 2010). Different effects of online reviews act on different stages of consumer decision-making in social commerce (Zhang and Benyoucef 2016). Online reviews, to a certain extent, can act as timely feedback of product or service quality. As a way of persuasion, different type of involvement consumers respond differently by online reviews (Lee et al. 2008; Park et al. 2007). Consumers with low-involvement are found more likely to be affected by online reviews, while consumers with high-involvement are more likely to be affected by the number of reviews only when the quality of negative comments is high (Park et al. 2007). Lee et al. (2008) tested the role of the number and proportion of negative reviews in product information processing and drew similar conclusions. As for those typically highly involved products such as digital cameras, Gu et al. (2012) found that online reviews in retailing websites had significant positive correlation with buying high-involvement products, but the degree was lower than that provided by other types of websites. We argue that the definitions of product involvement are anchored in the existing studies from a subjective perspective. Meanwhile how online reviews influence consumer involvement with those non-purchase decision remains mysterious.

Modeling Dynamic Shopping Cart Decisions with Heterogeneity

This RIP paper aims to examine the factors of determining virtual shopping cart choices in the process of interactive behaviors of using e-commerce websites. Simultaneous analysis of the purchase and non-purchase behaviors requires to measure SCA with actual objective data instead of surveyed or experimental data. ***Discarding cart*** typically refers that a customer places an item to the shopping cart, but leaves the website without a transaction of any item(s) (Ding et al. 2015; Kukar-Kinney and Close 2010; Li and Chatterjee 2006). We argue that this definition reflects the buying behavior which is unable to directly present a consumer's SCA behavior at the actual level of stock-keeping unit. This means although without checking out, the item still remains in the shopping cart and hence the abandonment is considered passive. Instead of focusing on such passive behavior and unknown final result (whether to check out when returning back to the website next time), in this study we aim to underline consumers' active behavior and thus define ***SCA behavior*** as a consumer's proactive remove of a stock-keeping-unit item from his/her virtual shopping cart. Based on this definition, we examine how an online consumer makes the choice of adding a specific item to shopping cart, and then either discards it or completes the transaction.

The period during which a user visits a website, including a series of interactive activities from entry to exit, is called a session in this study. Based on the literature (Ding et al. 2015; Li and Chatterjee 2006), the starting point of a complete session in this study refers to when a consumer enters the website and starts browsing its web page. The ending point is when the consumer stays on the same web page continuously for more than 20 minutes without making any new requests such as clicking a hyperlink. The ending point indicates the consumer leaves web pages untouched for more than 20 minutes. A session is the basic time interval unit in our calculation. For validating the choice of 20 minutes, robustness checks with different time intervals will be conducted.

In the subsection below, we construct a latent class model (LCM) to characterize consumer heterogeneity. Then we propose a discrete Choice Model (DCM) for modeling the dynamic virtual shopping cart decision at the stock-keeping-unit (SKU) level based on each individual consumer's clickstream data within a session.

Consumer Heterogeneity

Heterogeneity is a concept explained by the demographic variables of consumers, referring to different characteristics from different consumers instead of features of the market in the aggregate (Allenby and Rossi 1998; Chatterjee et al. 2003; Ding et al. 2015; Li and Chatterjee 2006). Considering that consumer heterogeneity is proximate to the nature of consumers, we use LCM to capture the heterogeneity (latent variables) based on consumers' demographic variables (manifested variables) to achieve consumer segment. Assuming that there is a latent variable "$Class$" whose values include $g(g=1,2,\cdots,G)$ classes, it explains the relationship between the different classes of consumers through ordered demographic variables. That is, gender D_1, age D_2, registration time duration D_3 and e-retailer membership level D_4 as shown in Table 1 in this study. The joint probability of LCM is

$$\pi_{ijkl}^{D_1D_2D_3D_4}=\sum_{g=1}^{G}\pi_g^{Class}\pi_{ig}^{Class(D_1)}\pi_{jg}^{Class(D_2)}\pi_{kg}^{Class(D_3)}\pi_{lg}^{Class(D_4)} \quad (1)$$

In formula (1) $\pi_{ijkl}^{D_1D_2D_3D_4}$ is the joint probability of G latent classes, representing the proportion of the combinations of the attributes' level $\{i,j,k,l\}$ of the total observed data. π_g^{Class} is the probability of latent class g and $\sum_{g=1}^{G}\pi_g^{Class}=1$.

Besides, $\pi_{ig}^{Class(D_1)}$ is the conditional probability of gender being i that a consumer belongs to the class g. Correspondingly, we denote $\pi_{jg}^{Class(D_2)}$, $\pi_{kg}^{Class(D_3)}$ and $\pi_{lg}^{Class(D_4)}$. LCM assumes that

$$\sum_i\pi_{ig}^{Class(D_1)}=\sum_j\pi_{jg}^{Class(D_2)}=\sum_k\pi_{kg}^{Class(D_3)}=\sum_l\pi_{lg}^{Class(D_4)}=1.$$

According to the Bayesian posterior probability criterion, each individual consumer is estimated the class she/he belongs to,

$$\hat{\pi}_{gijkl}^{ClassD_1D_2D_3D_4} = \frac{\hat{\pi}_{ijklg}^{D_1D_2D_3D_4Class}}{\sum_{g=1}^{G} \hat{\pi}_{ijklg}^{D_1D_2D_3D_4Class}} \quad (2)$$

$$\hat{\pi}_{ijklg}^{D_1D_2D_3D_4Class} = \pi^g \pi_i^{Class(D_1)} \pi_j^{Class(D_2)} \pi_k^{Class(D_3)} \pi_l^{Class(D_4)} \quad (3)$$

Wherein, $\hat{\pi}_{gijkl}^{ClassD_1D_2D_3D_4}$ is the posterior probability that an individual consumer belongs to class g. $\hat{\pi}_{ijklg}^{D_1D_2D_3D_4Class}$ is the probability that an individual consumer belongs to class g as well as demographic variables' values are $\{i, j, k, l\}$ correspondingly.

Virtual Shopping Cart Behavior

In reality, consumers cannot browse all products online. That means the number of various products is always sufficient online. Typically e-commerce platforms can only recommend a specific small number of products (from a large set of products) to their potential consumers as their consideration sets. In online shopping decision-making process, a consumer usually makes a lot of interaction with the website. The clicks in the shopping cart are those operations of consumers directly related to the purchase or non-purchase behavior (Ding et al. 2015; Li and Chatterjee 2006). Traditional choice models use market aggregate data, but the disadvantage is the neglect of individual characteristics of consumers. Considering each consumer can have his/her own preference toward products or brands, in this study we take individual consumers as the unit of analysis. We assume that consumption behavior is rational (Simon 1979). Decision-making outcomes are the responses of consumer rationalization according to their previous knowledge. At the time t of a session, the consumer i's potential utility U_{ijptg} of shopping cart decision to a specific product p within the G class formulated as below.

$$U_{ijpt} = \beta_{ijp0} + \beta_{ij1} I_{ijp(t-1)} + \beta_{ij2} R_{ijp(t-1)} + \beta_{ij3} B_{ijp(t-1)} + \varepsilon_{ijt} \quad (4)$$

Wherein, $I_{ijp(t-1)}$ is the product p's involvement vector accumulated to t-1 by consumer i with in a session. $R_{ijp(t-1)}$ represents an online review vector which the consumer i can view at the page of product p, whose mass online reviews information are accumulated to t-1. Similarly, $B_{ijp(t-1)}$ refers to a vector indicating the consumer i's past behavior, including $Buy_{ijp(t-1)}$, suggesting if consumer i has bought the product p before. $Weekend_{ijp(t-1)}$ refers to whether consumer i visits the focal e-commerce websites in weekends at t-1. The shopping cart choice preferences β_{ijp0} are captured by consumers. β_{ij1} is a parameter matrix of product involvement, and β_{ij2} is corresponding to online review preference and the past behavior parameter matrix is β_{ij3}. Random utility ε_{ijt} is independent and identically distributed in Gumbel distribution.

According to the principle of utility maximization, shopping cart decision is made by consumers based on the maximum potential utility. So we define the choice decision S_{ipt} of potential utility of consumer i, whose values are discrete $j \in \{1, 2, \cdots, C\}$, thus

$$S_{ipt} = \sum_{j=1}^{C} j \times I(max(U_{ipt}) = U_{ijpt} \quad (5)$$

$I(\cdot)$ is an indicator operator. When its value equals to 1, the current situation meets the requirements of the inner function; otherwise, its value is 0. We set up C as 4. If S_{ipt}=1, consumer i continues to browse the website but does not make a shopping cart choice. For S_{ipt}=2, consumer i places the product p in the shopping cart; For S_{ipt}=3, consumer i removes the product p from the shopping cart (i.e., abonnement). For S_{ipt}=4, consumer i purchases the product p. In order to ensure the estimation of the model, we use the general setting of the benchmark utility as 0. In equation (5), we set the consumer's behavior of browsing the website instead of making any choice decision about the shopping cart as benchmark (i.e., j=1 and U_{i1pt}=0). The probability is

$$P(S_{ipt}=1|I_{ijp(t-1)},R_{ijp(t-1)},B_{ijp(t-1)})=\frac{1}{1+\sum_{j=2}^{4}e^{\beta_{ij0}+\beta_{ij1}I_{ijp(t-1)}+\beta_{ij2}R_{ijp(t-1)}+\beta_{ij3}B_{ijp(t-1)}}} \quad (6)$$

Therefore, the probability of consumer i's choice behavior (i.e., $j\in\{2,3,4\}$) associated with the shopping cart related to a specific product p at the time t of a session is formulated as below.

$$P(S_{ipt}=j|I_{ijp(t-1)},R_{ijp(t-1)},B_{ijp(t-1)})=\frac{e^{\beta_{ij0}+\beta_{ij1}I_{ijp(t-1)}+\beta_{ij2}R_{ijp(t-1)}+\beta_{ij3}B_{ijp(t-1)}}}{1+\sum_{j=2}^{4}e^{\beta_{ij0}+\beta_{ij1}I_{ijp(t-1)}+\beta_{ij2}R_{ijp(t-1)}+\beta_{ij3}B_{ijp(t-1)}}} \quad (7)$$

The maximum likelihood estimation is used to estimate and solve the above parameters of the models. The shopping cost of consumers is not only reflected in price, but also manifested by consciously searching, evaluating and processing product information in e-commerce (Jiang et al. 2010; Schellong et al. 2016; Zaichkowsky 1986). From the perspective of implicit feedback of interactive behavior, the cumulative browsing time duration T_{ipt}, the cumulative number of clicks C_{ipt} and whether the product is tagged or not F_{ipt} , are three quantitate variables of the construct product involvement $I_{ijp(t)}$. Suppose that time duration of browsing a product page of consumer i at t-1 is T_{it}

$$T_{it}=Time_{i(t+1)}-Time_{it} \quad (8)$$

$$T_{ipt}=\sum_{t=1}^{t}T_{it}(p) \quad (9)$$

$Time_t$ represents the starting moment of the visit t. It declares the end of a session only when $T_{it}>20$ minutes.$T_{it}(p)$ is the subset of T_{it} , which is the time duration about the product p at t. Deduced by the formula (6), T_{ipt} is the browsing time duration accumulated to t by consumer i to the product p. The different preference of consumers leads to different cumulative browsing time for different products. So T_{ipt} is a measure of involvement horizon (Schellong et al. 2016), reflects the ***product-time involvement*** of processing product information for different consumers. The longer the T_{ipt} is, the higher consumer involvement in the product p, and the higher interested consumer i in it (Chatterjee et al. 2003;Schellong et al. 2016).

The statistic C_{ipt} , with a resemble definition and calculation to T_{ipt}, indicates the cumulative number of the product p clicks received up to t by consumer i. The ***product-clicking involvement*** is proxy of involvement frequency, and represents the amount of attention a consumer pays to a specific product over time (Schellong et al. 2016).

The relevance between the product and the consumer inner needs is one of production involvement measures. In the context of e-commerce, product tagging has a significant positive impact on the possibility to buy/sell a product (Ou and Chan 2014). Thus, the term whether consumer i has tagged the product p in past is constructed as the ***product-tagging involvement*** F_{ipt}

$$F_{ipt}=\begin{cases}1 & \text{If product p has been tagged to favorite set}\\ 0 & \text{Otherwise}\end{cases} \quad (7)$$

Online product reviews mainly provide information signals to potential consumers through the quantity and quality of reviews to make purchasing judgments (Gu et al. 2012; Lee et al. 2008; Park et al. 2007). This RIP paper also explores the impact of online reviews from these two aspects. We construct online reviews vector R_{ijpt} by the number of reviews nR_{ipt}, if a product has negative reviews $NegR_{ipt}$ and the proportion of negative reviews $RateNegR_{ipt}$. nR_{ipt} is the number of product p reviews at which consumer i can view at a time t.

$$nR_{ipt}=\begin{cases}0 & \text{No reviews}\\ 1 & \text{1-10 pieces of reviews}\\ 2 & \text{11-50 pieces of reviews}\\ 3 & \text{Over 50 pieces of reviews}\end{cases} \quad (8)$$

In this study, we will set online product reviews as a dummy variable for a better interpretability and $nR_{ipt}=0$ is operated as the benchmark. The variable is binary, referring to if a product has negative reviews $NegR_{ipt}$. When its value is 1, that means consumer i views product p which has negative reviews, at time t, or otherwise 0. Meanwhile, the proportion of negative reviews $RateNegR_{ipt}$ denotes the amount of the negative reviews versus the total number of reviews of product p at time t.

Data Collection and Analysis

The purpose of this RIP is to investigate the effects of product involvement and online reviews on consumers' virtual shopping cart decision. It requires a data set including the sequential browsing paths produced by the actual browsing behavior of individual-level consumers. Besides, the demographic information is vital for capturing consumer heterogeneity. Lastly, a data set from real-world e-retailer is considered necessary.

Table 1. Demographics of Consumers (n=78,746)

Variables	Mean	SD	Min	Max
Registration duration (days)	1088	722.62	1	3388
Gender	Male (42.98%), Female (7.34%), Classified (49.68)			
Age	below 15 (0.01%), 16-25 (8.19%), 26-35 (45.52%), 36-45 (28.98%), 46-55(3.09%), beyond 56 (1.77%), Classified (12.44%)			
Membership level	1 (1.86%), 2 (7.37%), 3 (21.48%), 4 (31.20%), 5 (38.09%)			

Notes: Gender, age and membership level are categorized variables. Some consumers chose not to disclose their gender or age information, and therefore labelled as "classified". "Registration duration" is a continuous variable. There are five levels of memberships. The higher the membership level is, the more amount of consumption has spent in the focal platform.

Based on the above research objectives, we acquired a real clickstream data set from a large Chinese e-commerce platform. This data set contains users' web logs in a recent day. Each single data entry represents as a specific webpage as begun when the user started the interaction with the webpage. The raw data set consists of 78,746 distinct consumers and 19,508 stock-keeping-unit products among 399 brands in 8 product categories. The frequency of human-computer-interactions (HCI) is 13,199,934 action points, including clicking, tagging into a favorite set, adding to shopping cart, discarding product(s) from shopping cart (i.e., SCA) and purchase. The demographic characteristics of consumers is summarized in Table 1.The calculation of the above formulas will be conducted in R language.

Conclusions and Expected Contributions

Prior literature focuses on either purchasing behavior (Gu et al. 2012; Kim et al. 2004; Mudambi and Schuff 2010; Ou and Chan 2014) or SCA behavior (Berry et al. 2002; Ding et al. 2015; Kukar-Kinney and Close 2010). The contributions of this RIP paper is mainly embodied in the theoretical integration of purchase and non-purchase behavior in online decision-making process. We define the SCA behavior as an online consumer's proactive remove of stock-keeping-unit item(s) from their shopping cart instead of leaving the website without a transaction of any items. We exploit a latent class model (LCM) and a discrete choice model (DCM) to model the dynamic virtual shopping cart decision with the consideration of consumer heterogeneity. This study provides a novel method to analyze consumer behaviors in view of a large set of individual-level clickstream data. Besides, we fractionize the consumer-website interaction specific product involvement, in an attempt to extend the literature on consumer involvement to the disciplines of information systems and human-computer interaction by analyzing and managing the large clickstream data. Last, but not the least, our proposed models cover the effects of online reviews and consumer involvement and thus extend the extant literature on virtual shopping abandonment to a relative new context of social commerce. We also expect the results of this study will benefit online retail platforms to better securitize the role of product involvement based on consumers' implicit needs, and thus personalize their marketing strategies and managing online reviews information in a better form.

Acknowledgements

We acknowledge funding support from the National Science Foundation of China # 71671048.

References

Allenby, G. M., and Rossi, P. E. 1998. "Marketing models of consumer heterogeneity," *Journal of Econometrics* (89:1-2), pp. 57-78.

Berry, L. L., Seiders, K., and Grewal, D. 2002. "Understanding service convenience," *Journal of Marketing* (66:3), pp. 1-17.

Bian, X., and Moutinho, L. 2011. "The role of brand image, product involvement, and knowledge in explaining consumer purchase behaviour of counterfeits: Direct and indirect effects," *European Journal of Marketing* (45:1/2), pp. 191-216.

Bucklin, R. E., and Sismeiro, C. 2009. "Click here for internet insight: advances in clickstream data analysis in marketing," *Journal of Interactive Marketing* (23:1), pp. 35-48.

Chatterjee, P., Hoffman, D. L., and Novak, T. P. 2003. "Modeling the clickstream: Implications for web-based advertising efforts," *Marketing Science* (22:4), pp. 520-541.

Detlor, B., Sproule, S., and Gupta, C. 2003. "Pre-purchase online information seeking: Search versus browse," *Journal of Electronic Commerce Research* (4:2), pp. 72-84.

Ding, A. W., Li, S., and Chatterjee, P. 2015. "Learning user real-time intent for optimal dynamic web page transformation," *Information Systems Research* (26:2), pp. 339-359.

Gu, B.,Park, J.,and Konana, P. 2012. "Research note: the impact of external word-of-mouth sources on retailer sales of high-involvement products,"*Information Systems Research* (23:1), pp.182-196.

Huang, L., Tan, C. H., Ke, W., and Wei, K. K. 2018. "Helpfulness of Online Review Content: The Moderating Effects of Temporal and Social Cues," *Journal of the Association for Information Systems* (19:6), pp. 503-522.

Jiang, Z., Chan, J., Tan, B. C., and Chua, W. S. 2010. "Effects of interactivity on website involvement and purchase intention," *Journal of the Association for Information Systems* (11:1), pp.34-59.

Kim, H.-W., Xu, Y.,and Koh, J. 2004. "A comparison of online trust building factors between potential customers and repeat customers,"*Journal of the Association for Information Systems* (5:10), pp. 392-420.

Kukar-Kinney, M., and Close, A. G. 2010. "The determinants of consumers' online shopp.ing cart abandonment," *Journal of the Academy of Marketing Science* (38:2), pp. 240-250.

Lee, J., Park, D.-H., and Han, I. 2008. "The effect of negative online consumer reviews on product attitude: An information processing view," *Electronic Commerce Research and App.lications* (7:3), pp. 341-352.

Li, S., and Chatterjee, P. Year. 2006 "Shopp.ing cart abandonment at retail websites-a multi-stage model of online shopp.ing behavior," *IEEE International Conference on Data Mining*, Springer.

Mudambi, S. M., and Schuff, D. 2010. "Research note: What makes a helpful online review? A study of customer reviews on Amazon. com," *MIS Quarterly* (34:1), pp. 185-200.

Ou, C. X., and Chan, K. C. 2014. "Developing a competitive edge in electronic markets via institutional and social based quality signaling mechanisms,"*Information & Management* (51:5), pp. 532-540.

Park, D.-H., Lee, J., and Han, I. 2007. "The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement," *International Journal of Electronic Commerce* (11:4), pp. 125-148.

Richins, M. L., and Bloch, P. H. 1991. "Post-purchase product satisfaction: Incorporating the effects of involvement and time," *Journal of Business Research* (23:2), pp. 145-158.

Schellong, D., Kemper, J., and Brettel, M. Year. "Clickstream data as a source to uncover consumer shopp.ing types in a large-scale online setting," *The proceeding of the 24th European Conference on Information Systems, ECIS 2016,* Istanbul, Turkey.

Simon, H. A. 1979. "Rational decision making in business organizations," *The American Economic Review* (69:4), pp. 493-513.

Zaichkowsky, J. L. 1986. "Conceptualizing involvement," *Journal of Advertising* (15:2), pp. 4-34.

Zhang, K. Z. K., and Benyoucef, M. 2016. "Consumer behavior in social commerce: A literature review," *Decision Support Systems* (86:6), pp. 95-108.