Identifying market structure to monitor product competition using a consumer-behavior-based intelligence model

Mingjun Zhan
School of Management, Guangdong University of Technology, Guangzhou, China
Hongming Gao
School of Management, Guangdong University of Technology, Guangzhou, China and Faculty of Business and Economics, The University of Hong Kong, Pokfulam, China
Hongwei Liu, Yidan Peng and Dan Lu
School of Management, Guangdong University of Technology, Guangzhou, China, and
Hui Zhu
Department of Electronic Commerce, Guangzhou University, Guangzhou, China

Received 20 August 2019
Revised 9 December 2019
30 December 2019
Accepted 20 January 2020

Abstract

Purpose – The objective of this paper is to propose a consumer-behavior-based intelligence (CBBI) model to identify market structure so as to monitor product competition. Competitive intelligence extracted from Chinese e-business clickstream data is exploited to examine the relevance of consumers' heterogeneous behavioral feedback, namely, click, tag-into-favorite, time-of-browsing, add-into-cart, and remove-from-cart, to visualize the competitive product market structure and to predict product-level sales.

Design/methodology/approach – Our proposed CBBI model consists of visualization and prediction, which explore e-business clickstream data. We conduct the visualization and segmentation of market structure in the form of a perceptual map by employing K-means clustering algorithm and multidimensional scaling technique. Concurrently, we developed an updated Bayesian linear regression (BLR) to predict product-level sales by considering consumers' heterogeneous feedback. Our updated BLR specifically integrated the estimated knowledge of the previous periods to verify whether product sales are period-dependent due to the consumer memory effect in e-commerce, improving the conventional BLR of diffuse prior distribution setup in terms of mean absolute error (MAE) and root mean squared error (RMSE).

Findings – Considering the performance of consumers' heterogeneous actions, the present research visualized three different segments of the competitive market structure in a perceptual map, and its horizontal axis is shown as a signal of the ascending trend of product sales. The previous five-day period was ascertained to be the best size of a time window for the consumer memory effect on product sales prediction. This hypothesis is supported by the concept that product sales are period-dependent. The results of the proposed updated BLR indicate that consumer tag-into-favorite, add-into-cart, and remove-from-cart feedback have positive impacts on product-level sales while click and time-of-browsing have the opposite effect.

Originality/value – While the identified competitive product market structure elaborates consumer heterogeneous feedback toward alternative product choices, this paper contributes by extending those homogeneous consumer preferences-related marketing studies. The perceptual map's configuration in respect to period-dependent product sales facilitates the effective inclusion of consumer behavior application in product sales prediction research in e-commerce. This paper helps sellers and retailers better comprehend the impacts of heterogeneous feedback and the consumer memory effect on the degree of competition in the form of

This research was supported by the National Natural Science Foundation of China [grant number 71671048]; Guangdong Construction of High-Level Colleges for Postgraduate Study Abroad Project in Guangdong University of Technology 262515006 [grant number 262515006]; and Top Innovation Graduate Student Cultivation Project Fund.

Asia Pacific Journal of Marketing and Logistics

1355-5855
DOI 10.1108/APJML-08-2019-0497

product sales. The research results also offer a managerial implication about shaping the competitive edge by conducting different product management strategies in e-commerce platforms.

Keywords Market structure, Competitive intelligence, Product competition, Product sales, Consumer memory effect, Clickstream, Bayesian linear regression, Perceptual maps

Paper type Research paper

1. Introduction

The competition to Amazon.com, the leader of global electronic commerce marketplaces, experienced an 8 percent increase from 2017 to 2018 (Grasso, 2018). More and more online sellers struggle to constantly adjust their product management strategies to cope with intensive market competition. Specifically, firms in e-business tend to choose their products by analyzing market structure, speculating what is popular in a certain marketplace, and what their competitors are selling. It is well recognized that consumer feedback or market responses expedite knowledge management for competitive product market structure, which is the core strategy for helping firms shape their competitive advantage (Day *et al.*, 1979; Porter, 1979; Phau and Lum, 2000). However, in the current dynamic competition environment, it requires real-time data and effective business models or technological advancement in the discipline of marketing to conduct accurate market segment strategies to achieve dynamic product positioning.

To positively influence product competition, researchers on market structure have made efforts to mitigate adverse consequences caused by intensive competition. Current survey-based data research (DeSarbo *et al.*, 2006; DeSarbo and Grewal, 2007) and scanning panel data investigation (Erdem, 1996; Leszczyc *et al.*, 2000) constitute consumer judgmental studies. However, the former is constrained by consumer recall bias and the effect of uncertain demand (Netzer *et al.*, 2012), while the latter disregards the consideration set of alternative product choices in the prepurchase stage during a whole shopping process. Employing the data source of consumer judgmental studies, however, makes it difficult to completely understand the dynamics of a competitive product market.

Thanks to the development of the information technologies of Web 2.0, a variety of user activity-related data is available for product market analysis. Human–computer interaction and interface design are now critical factors for a successful retail business (Ijaz *et al.*, 2016). User experiences of shopping are affected by various factors, such as time consumed, products purchased, the level of fatigue involved, and so on. It has also been found that interactive displays can deliver product information in an effective way, which can enhance shopping experiences and sales (Ijaz *et al.*, 2014; Ijaz and Rhee, 2018). The degree of user interaction based on consumer behavior online depends on a variety of Internet marketing strategies (Wu, 2002). Some researchers have scrutinized competitive product market structure, through such factors as frequency of products in online reviews (Bao *et al.*, 2008; Lee and Bradlow, 2011; Netzer *et al.*, 2012), search keywords (Wei *et al.*, 2016), or weblog data of e-commerce (Kim *et al.*, 2011; Ringel and Skiera, 2016). These previously mentioned researches have verified that competitive product market structure could be measured well by online users' behavioral feedback. However, they focus on a homogeneous measurement that manifests consumer preferences of products. Investigation into the effects of consumer's heterogeneous feedback on product competition and how to build a product-level or brand-level competitive edge in product market structure remains underexplored. Furthermore, prior research has primarily considered the local competitive level among products and their choice substitutes (DeSarbo *et al.*, 2006; DeSarbo and Grewal, 2007; Bao *et al.*, 2008). However, product sales volume is one of the most explicit measurements of the global competitive level in the electronic market (Ou and Chan, 2014; Ringel and Skiera, 2016). Intuitively, the higher sales a product owns, the stronger its competitive edge becomes. Thus, we can quantify the product competition in the form of product-level sales.

In the context of e-commerce, consumer feedback takes various forms. For example, a consumer can click on an item page implying her preference toward the focal product (Moe and Fader, 2004; Ding *et al.*, 2015). Also, the time that she spends on browsing the item detail page might help infer whether she intends to purchase or just browse (Montgomery *et al.*, 2004; Su and Chen, 2015; Raphaeli *et al.*, 2017). In the case that she tags the item to favorites (e.g. the different forms of favorites: the List in Carrefour, France; the Wish List in Amazon, USA; the Shoucangjia in Taobao, China), this shows a positive signal of the likelihood that she will buy the item (Ou and Chan, 2014). Moreover, changes in deletions and additions of product substitutes within a shopping cart might be vital for determining competitive product market structure (Day *et al.*, 1979; Kukar-Kinney and Close, 2010), but these changes are still not a concern in product market research. Hence, this study argues that consumer click, time-of-browsing, tag-into-favorite, add-into-cart, and remove-from-cart are valuable for identifying the competitive product market structure and predicting product-level sales to monitor product competition.

Prior historical knowledge could influence consequent rational decision-making (Simon, 1979), owing to the consumer memory effect (Yan *et al.*, 2009; Strandburg, 2013). Akin to sales prediction, demand forecast is found to be more accurate as a result of a state-dependent policy (Azoury and Miyaoka, 2009), verifying the consumer memory effect in e-commerce. In our study, we found that product sales might be period-dependent as consumers browse and tag products into favorites in the current period but which they plan to buy later. We argue that the current consumer decision is bounded rational as it is influenced by their memory effect, that is, historical knowledge and previous experiences (Simon, 1979; Yan *et al.*, 2009; Strandburg, 2013). Therefore, we hypothesize that product sales are period-dependent. In essence, the estimated knowledge from the previous period significantly facilitates the accurate prediction of product sales.

To fill in this research gap about consumer's heterogeneous feedback and the memory effect of rational decision-making in the literature of market structure, this work aims to examine the potential impacts of consumers' heterogeneous behavioral feedback and consumer memory effect on product competition using real-time clickstream data. Specifically, this study investigates the following:

(1) Are product sales period-dependent? Does the estimated knowledge from the previous period facilitate the accuracy of product sales forecast due to the consumer memory effect?

(2) Can the consumers' heterogeneous feedback, such as click, time-of-browsing, tag-into-favorite, add-into-cart, remove-from-cart, be exploited to visualize and identify competitive product-level market structure?

(3) What are the effects and valence of consumers' heterogeneous feedback on predicting the degree of product competition in the form of product-level sales?

Hence, this study proposes a consumer-behavior-based intelligence (CBBI) model to monitor product competition by visualizing product-sale market structure in a perceptual map. Based on consumer intelligence about users and objectiveness in clickstream data from one of the top e-commerce platforms in China, our proposed CBBI model includes the visualization part describing product market structure and the prediction part predicting product sales. Firstly, similar to the past research (Kim *et al.*, 2011; Ringel and Skiera, 2016), we apply the conventional K-means clustering algorithm and multidimensional scaling (MDS) technique to approach market segmentation and visualization. Secondly, we propose updated Bayesian linear regression (BLR) for product sale prediction. There are several advantages of employing updated BLR in our study. The BLR is appropriate in a high-uncertainty environment of product sale forecasting (Azoury and Miyaoka, 2009). Specifically, point estimation in ordinary least squares (OLS) estimates the unknown value for a certain parameter of independent variable too precisely, resulting in overfit issue. The BLR relaxes this restriction and involves uncertainty by outputting posterior probability distribution (PPD) (Rossi and Allenby, 2003; Albert, 2009). In addition, our proposed updated BLR is a dynamic approach, which is different from stable linear regression (Rossi and Allenby, 2003; Albert, 2009). We incorporate the prior historical knowledge to update the product sale data in the current period since the product sales' high uncertainty and the consumer memory effect can then be determined. Overall, the proposed updated BLR can decrease the product sales' uncertainty and learn the consumer memory effect by allowing for prior historical knowledge and outputting predictive distribution. Last, but not least, as different durations of the memory effect could influence the accuracy of product sales prediction, we conduct a sliding window approach in our proposed updated BLR to ascertain the best size of a time window endogenously based on the data.

The identified competitive product market structure elaborates consumer heterogeneous feedback toward alternative product choices, and this study theoretically extends research on homogeneous feedback for product market structure (Kim *et al.*, 2011; Netzer *et al.*, 2012; Ringel and Skiera, 2016). At the same time, this study contributes to the existing studies on consumer memory effect and product competition. In terms of managerial implications, the findings may help sellers and retailers more deeply understand competitive product market structure with respect to consumer behaviors, as well as offer a theoretical backup to shape their competitive edge by conducting different product management strategies in e-commerce platforms.

2. Related work

2.1 Product market and traditional methods

Competitive product market structure is defined as the set of products that consumers determine to be substitutes within relevant specific usage situations (Day *et al.*, 1979). Elrod *et al.* (2002) considered consumer influences and defined market structure analysis, which aims at obtaining a fundamental understanding of competition, broadly as explanations of the extent to which the products/services under consideration are substitutes or complements. Managers can benefit from market structure insight and make better market strategies for not only the design and development of new products but also repositioning of the existing products (Lee and Bradlow, 2011). For a certain product category, market structure can be characterized as a set of product submarkets for useful representation of competition. That is, these submarkets reflect the discrete representation of competition behind the entire market structure. Previous studies showed several ways of defining a submarket partition, including a combination of features or consumers' perceptions of quality and style (Urban *et al.*, 1984; France and Ghose, 2016) and a combination of cultural, individual, and institutional sources (Glushko *et al.*, 2008). Moreover, a submarket can be defined as a brand at the lowest level of granularity. However, complete brand loyalty is not always assumed for every product category if each brand is considered as separate submarket (France and Ghose, 2016). Ehrenberg *et al.* (2004) found that a great degree of substitutability among brands occurs in the frequently purchased consumer product markets. Thus, the relationship among brands should be considered.

A large number of methodologies pervade market structure analyses, depending upon either judgmental or behavioral data. Among traditional data analyses, customer judgmental studies can reveal the dynamics of consumer-based cognitive product marketing. DeSarbo *et al.* conducted a survey about automobiles and presented competitive asymmetric characters of product marketing (DeSarbo *et al.*, 2006; DeSarbo and Grewal, 2007). Generally, since consumers are bounded by cognitive capacity, survey-based data commonly indicate that consumers might answer ambiguously owing to memory recall bias, or unexpected demand in the future, namely, the uncertain demand effect (Netzer *et al.*, 2012). This effect produces higher uncertainty about whether consumers can recall their latest purchase decisions and if this affects their intentions to purchase in the future.

Another traditional stream of literature relies on scanner panel data to infer the intention hidden in consumers' purchase behavior (Erdem, 1996; Leszczyc *et al.*, 2000). For instance, Erdem (1996) devised a market structure among brands of margarine and liquid detergent by modeling consumer purchased goods in shopping baskets. Also, Leszczyc *et al.* (2000) modeled scanner panel data to examine the market structure of grocery stores. Nonetheless, one of the disadvantages of panel data is that such analysis commonly requires repeat transactions of consumables without considering durables (Ringel and Skiera, 2016). What is more, it merely concentrates on products in individual consumers' shopping baskets instead of all products. That is, scanner panel data can only serve to investigate the product market in the postpurchase phase, which disregards the consideration set in the prepurchase stage.

2.2 Previous analysis methodologies to identify market structure

Recently, improved information technologies have been effectively developed in Web 2.0. Enormous amounts of online Web information have attracted scholars' attention concerning competitive product market structure identification based on either well-structured or unstructured data.

The consumers' feedback in discussing text is one flexible source to identify competitive product structure and discover primary competitors of a business entity. Bao *et al.* (2008) created a Web information extraction method for identifying main competitors based on the content discussed and compared by Web users. Acquiring co-occurrence of brands in messages as a key variable using text mining methods, Netzer *et al.* (2012) provided competitive landscape of sedan cars and diabetes drugs from online forums. In addition, Lee and Bradlow (2011) analyzed structured pros and cons lists (i.e. phrases) to generate product attribute market structure, based on finer granularity of online reviews at a review platform. They found that online herding bias is concealed in online opinions, for users attempt to have views liked popularly and that can be accessed publicly.

Online search data offer another perspective of consumers' feedback in a competitive product market. To cope with the great heterogeneous of consumer behavior and preference, there is a need for market segmentation from the perspective of marketing management. (Bruwer *et al.*, 2017; Su and Liu, 2017; Sultan *et al.*, 2018). Research could also benefit from a facilitating tool known as mapping in a more visual representation, which delineates where things are in relation to one another intuitively (MacInnis, 2011). Kim *et al.* (2011) proposed a multiattribute method to visualize competitive camcorder market structure with product search data from Amazon.com. They utilized homogeneous feedback, users' searching behavior. Products searched in the same specific period were identified as competitive products, and they found that such competitive networks are verified to be effective by daily percentile rankings of products. Ringel and Skiera (2016) similarly used the single consumer search action in the LED-TV market as input for visualization in low-dimensional mapping. Wei *et al.* (2016) exploited search keywords to model a novel bipartite graph, specifically quantified the local competitive degrees among brands in the fields of computers and cars. They found that conjoint keywords and their search volume from query logs have a significant influence on forecasting market share monthly.

These previous research findings have greatly facilitated competitive product market structure analysis relying on online search data. Nonetheless, the current studies focus on the homogeneous feedback, which indicates the co-occurrence of products displaying a local competitive edge (Bao *et al.*, 2008; Lee and Bradlow, 2011; Netzer *et al.*, 2012; Ringel and Skiera, 2016; Wei *et al.*, 2016).

2.3 Business intelligence analysis using big clickstream data

Clickstream data capture individual users' activities online from weblogs (Bucklin and Sismeiro, 2009). Within the context of e-commerce, the clickstream data are used to investigate site usage and consumer browsing behavior (Moe and Fader, 2004; Montgomery *et al.*, 2004; Su and Chen, 2015; Raphaeli *et al.*, 2017), recommendation systems and purchase behavior (Sismeiro and Bucklin, 2004; Olbrich and Holsing, 2011; Nishimura *et al.*, 2018; Alfian, 2019), and persuasion and advertising effects on the Internet (Ansari and Mela, 2003; Bonfrer and Drèze, 2009; Rutz and Bucklin, 2012).

The sequence of pages visited by individual-level consumers previously motivated scholars to consider consumer visiting patterns online for webpage optimization and website design (Moe and Fader, 2004; Montgomery *et al.*, 2004). Clickstream data provide the opportunity to dynamically capture cumulative browsing effects consistent with real consumer behavior. Su and Chen (2015) considered these accumulating visit behaviors as information that comprehensively reflects diverse consumer interest patterns. Raphaeli *et al.* (2017) studied consumer intent orientation in mobile service and PC devices by their navigation sequence across different platforms.

Moreover, purchase conversion research appeals to e-commerce managers. In the application of clickstream, the probability of Internet buying is mostly estimated by either individual consumer purchase decisions, such as a binary probit model (Sismeiro and Bucklin, 2004), or segments of aggregated consumers, such as a latent class model (Nishimura *et al.*, 2018). In social commerce, Olbrich and Holsing (2011) investigated interface features and social shopping factors for predicting purchase likelihood of consumers depending on clickstream. Certainly, the most direct application of clickstream data in purchase conversion is recommendation systems (Bucklin and Sismeiro, 2009). Alfian (2019) exploited similar customers' browsing history and transactional data to achieve recommendation in the case of digital signage marketing.

Click-through rate is an intuitive standard metric for evaluating the performance of persuasion and advertising online. Ansari and Mela (2003) proposed a two-phase method to investigate the effects of e-mail customization on click-through rate. Based on the temporality of clickstream data, Bonfrer and Drèze (2009) forecasted and monitored the marketing performance of e-mail campaigns in real time. Rutz and Bucklin (2012) found that consumers' next choices of brand-specific pages to view are influenced by banner ad exposure in their current browsing session.

These informative studies have been driven by the details of the individual-level granularity, temporality, and cumulative effect hidden in clickstream data. They have provided an opening viewpoint for exploring market structure in recent times, as well. Market segmentation of consumers is identified by personality characteristics classified into small-scale clickstream data (Wen and Peng, 2002). But, few studies have focused on the clickstream within the product market. However, some scholars collected big clickstream data to analyze camcorder and LED television market structures, but as mentioned previously, the studies (Kim *et al.*, 2011; Ringel and Skiera, 2016) considered the competitive edge of products in the single search data instead of consumer deletion and addition feedback toward alternative product choices in the course of a shopping journey.

3. The CBBI model

Unlike a single measurement that models local competitive ratios among products in most existing studies, this study proposes a CBBI model for modeling consumer heterogeneous feedback from sequential individual users' activities online. This CBBI model is a business analysis model that could predict future sales and visualize the market structure depending on product category itself. Moreover, it considers general consumer heterogeneous feedback in global e-commerce platforms, such as click, tag-into-favorite, time-of-browsing, add-into-cart, and remove-from-cart, to visualize the competitive product market structure and to predict product-level sales. Take tag-into-favorite, for example. There are different forms of favorites, such as the Listes in Carrefour, France, the Wish List in Amazon, USA, and the Shoucangjia in Taobao, China. In sum, the CBBI model is convenient and appropriate for global e-commerce. A conceptual workflow is illustrated in Figure 1.

Through human–computer interactive feedback between service and client sides (e.g. mobile phones or computers), the electronic marketplaces record all consumers' actions from server sides in a B2B/B2C environment. It is the clickstream data recording consumer transaction information and interactive feedback that the CBBI model applies to visualize the focal market structure and predict product-level sales. The CBBI model consists of an intuitive *visualization* part and a *prediction* part. The product competition identified by CBBI might provide an efficient and low-risk tactic for relevant stakeholders, especially for sellers, supply chain managers, and manufacturers.

The purpose of competitive product market analysis is designed to be much more of a description than a prediction (Netzer *et al.*, 2012). For straightforward description of product competition, achieving market segmentation and visualization of product or brand competitors is vital (Kim *et al.*, 2011; Ringel and Skiera, 2016). Therefore, in the visualization part of CBBI, we employed the conventional K-means clustering algorithm

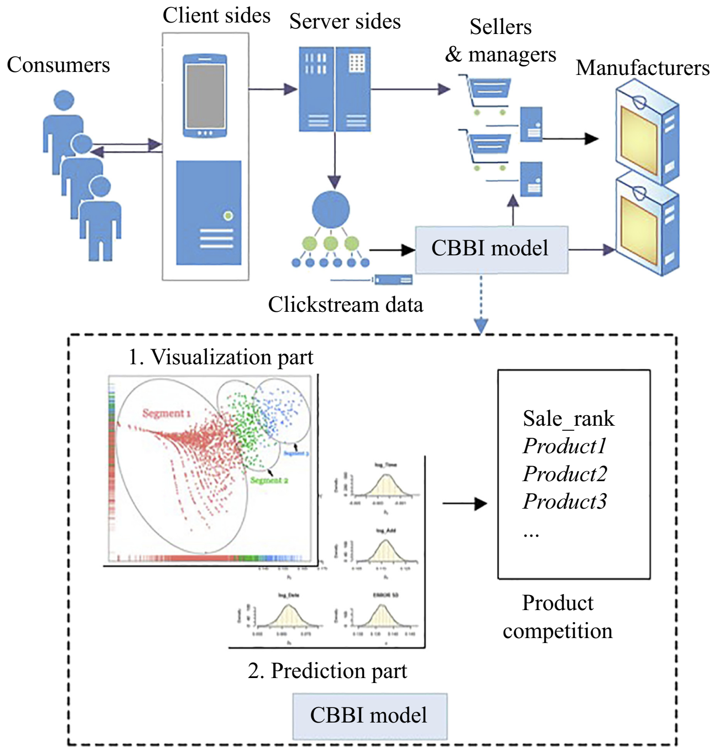


Figure 1.
Consumer-behavioral-based intelligence model

and MDS technique in the field of product market structure (Kim *et al.*, 2011; Ringel and Skiera, 2016). Also, two evaluation indicators were used to determine the optimal number of segments in the focal market structure. The silhouette index was used to interpret and verify the consistency of the clusters theoretically (Savaresi and Boley, 2004), while the heuristic Elbow principle was applied to ascertain the practicability (Kodinariya and Makwana, 2013). The visualization part for a description of product market structure is demonstrated in Section 5.

3.1 Updated BLR model

The BLR model embraces probability distribution to conduct linear regression rather than point estimation in OLS. The purpose of BLR is not to determine the single best value of parameters but to determine the probability posterior distribution (Rossi and Allenby, 2003; Albert, 2009). There are two strengths of the BLR. It can not only adapt prior knowledge (prior distribution) from experts in the model but also quantify the uncertainty of the model due to its output of probability distribution. Conventionally, BLR utilizes the conventional diffuse prior distribution setup (noninformative prior distribution), assuming estimation merely depending on data. However, similar to sales predication, Azoury and Miyaoka (2009) exploited BLR and found that the state-dependent demand forecast policy is optimal. In other words, it is sensible to assume that the current consumer decision is bounded rational as it is impacted by the consumer's historical knowledge and previous experience (Simon, 1979). Hence, we assume that product sale is period-dependent, since consumers might browse and tag products in the current period, but they could buy later. Thus, BLR was suitable for us to examine our hypothesis.

We set Sale_{it} as sale of a product i in the period t, observing n individual products in total m time periods, $t \in \{w_1, w_2, \ldots, w_m\}$. The normal regression model could be expressed as

$$\text{Sale}_{it} = \beta_0 + \beta_1 \text{Click}_{it} + \beta_2 \text{Time}_{it} + \beta_3 \text{Tag}_{it} + \beta_4 \text{Add}_{it} + \beta_5 \text{Rem}_{it} + \varepsilon_{it}$$

$$\varepsilon = (\varepsilon_{1t}, \ldots, \varepsilon_{nt}) \overset{iid}{\sim} N\left(0, \sigma^2\right), \;\; i = 1, 2, \ldots, n \tag{1}$$

or expressed specifically as

$$\text{Sale}_t \,\middle|\, \beta, \sigma^2, X_t \sim N_n\left(X_t\beta, \sigma^2 I_n\right) \tag{2}$$

where X_t is defined as the heterogeneous behavioral feedback 5×1 vector in the current period t. The corresponding parameter vector is set as $\beta = (\beta_1, \beta_2, \cdots, \beta_5)^{'}$, and I_n is an identity matrix. Also, $N_n(\mu = X_t\beta, A = \sigma^2 I_n)$ is n-dimension normal distribution, μ is a mean vector, and A is a covariance matrix. The error vector ε follows as a standard normal distribution, and $\varepsilon_{it}(1 \leq i \leq n)$ is independent and identically distributed.

This study defines heterogeneous behavioral feedback X_t in a specific period t as follows. We define $\text{Click}_{it} = \sum_{k \in K} \text{Click}_{kit}$ as the frequency of a focal product i clicked and then viewed by all K individual consumers. Simultaneously, the accumulated time that consumers K have browsed product i is represented as $\text{Time}_{it} = \sum_{k \in K} \text{Time}_{kit}$. In the context of e-commerce, the behaviors of product tagging and adding into a shopping cart have a significant positive impact on the possibility of the consumer buying a product (Ou and Chan, 2014). We argue that two types of feedback, tag-into-favorite and add-into-cart, are associated with product sales. Thus, the frequencies of a focal product i tagged into a favorite set and added into a shopping cart are denoted as $\text{Tag}_{it} = \sum_{k \in K} \text{Tag}_{kit}$ and $\text{Add}_{it} = \sum_{k \in K} \text{Add}_{kit}$, respectively. Moreover, consumer preferences, such as consumer deletion behavior, need to be involved in the analysis of product marketing (Day *et al.*, 1979; Ding *et al.*, 2015). Then, we calculate the frequency of removing a focal product i from shopping carts by all K consumers as $\text{Rem}_{it} = \sum_{k \in K} \text{Rem}_{kit}$.

In the ordinary linear regression model, variances σ^2 are assumed to be equal. Instead, the Bayesian inference approach commonly sets $\theta = (\beta, \sigma^2)$ as an unknown parameter vector. As in the study of Zellner (1976), we use diffuse prior distribution during the first period t:

$$p(\beta_t, \sigma_t^2) \propto (\sigma_t^2)^{-1}. \tag{3}$$

In the light of the properties of multivariate normal distribution (Zellner, 1976; Albert, 2009), the posterior distribution by Bayesian inference of the model can be inferred as:

$$p(\beta_t, \sigma_t^2 \mid \text{Sale}_t) = p(\sigma_t^2 \mid \text{Sale}_t) p(\beta_t \mid \sigma_t^2, \text{Sale}_t) \tag{4}$$

$$\beta_t \mid \sigma_t^2, \text{Sale}_t \sim N_n(\widehat{\beta}_t, V_{\beta_t} \sigma_t^2) \tag{5}$$

$$\sigma_t^2 \mid \text{Sale}_t \sim \text{invgamma}((n-k)/2, S/2) \cdot \tag{6}$$

Accordingly, we utilize stochastic simulation to sample σ_t^2 first by $p(\sigma_t^2|\text{Sale}_t)$ and then β_t by $p(\beta_t|\sigma_t^2, \text{Sale}_t)$, which could estimate the posterior distribution of parameters $\widehat{\theta}_t = (\widehat{\beta}_t, \widehat{\sigma}_t^2)$. The estimated parameters $\widehat{\theta}_t = (\widehat{\beta}_t, \widehat{\sigma}_t^2)$ in period t keep constant to forecast product sales in the consequent period, and this remains typical of noninformative prior BLR (Albert, 2009; Zellner, 1976). In this study, we take it as the stable BLR.

As the BLR can involve uncertainty by incorporating prior historical knowledge (Rossi and Allenby, 2003; Albert, 2009), it is a dynamic approach different from stable linear regression (Azoury and Miyaoka, 2009). Hence, we incorporate the use of prior historical knowledge to update the product sales in the current period, since the high uncertainty of product sales could be learned by BLR. Generally, we update the prior information with the average of posterior distribution $\widehat{\theta}_t = (\widehat{\beta}_t, \widehat{\sigma}_t^2)$ in the previous period t instead of diffuse prior distribution in a stable BLR. Also, we compare the performance of predicting the future product sales between the stable BLR and our proposed updated BLR to test the hypothesis that the product sales are period-dependent.

In addition, considering consumer memory effect as a marketing strategy, Yan *et al.* (2009) used one day and seven days as the short-term and long-term period, respectively, and concluded that targeting user behavior within a short-term period is more effective than within a long-term period. Based on their research, Strandburg (2013) suggested that such short-term tracking on advertising is contextually based. When memory effect involves the product market structure, except for the period-dependent assumption, the specific duration for the memory effect is not given endogenously, and it needs to be exogenously determined by the data. Empirically, ascertaining the best size of the time window is essential for preparing a more appropriate management decision earlier to avoid deteriorative conditions in the future.

Therefore, we propose an updated BLR model that conducts a sliding window approach (Chu, 1995; Deypir *et al.*, 2012). Such an approach does not only incorporate the prior knowledge for clickstream data but also identifies the best time period owing to its emphasis on the current period and its bounded memory requirement (Simon, 1979).

In the proposed updated BLR model, we divide our sequential data into m time periods $w = \{w_1, w_2, \ldots, w_m\}$ (i.e. m time windows) in a manner of a sliding window approach shown in Figure 2, wherein the size of the time window is $\|w_j\| = \lambda$ and $j = 1, 2, \ldots, m$. The issue of ascertaining the best size of the time window is transformed to solve λ^* exogenously when we optimize the performance of the updated BLR. That is to say, λ is a parameter defined by exogeneity. Specifically, we infer $\widehat{\theta}_{w_1} = (\widehat{\beta}_{w_1}, \widehat{\sigma}_{w_1}^2)$ in the first period w_1 from the stable BLR method. Recursively, the prior information of $\theta_{w_{j+1}} = (\beta_{w_{j+1}}, \sigma^2_{w_{j+1}})$ in the current period w_{j+1} is updated by the estimated mean of $\widehat{\theta}_{w_1} = (\widehat{\beta}_{w_1}, \widehat{\sigma}^2_{w_1})$ in the previous window w_j. For sake of the expression of simplicity and consistency with Eqns 1–6, it is worth mentioning the setup $t \in \{w_1, w_2, \ldots, w_m\}$ in the following equations.

Given predictor $X_t = x_t^*$ by the knowledge of multivariate Bayesian model, the posterior predictive distribution of product sale Sale_t will be:

$$p\left(\tilde{\text{Sale}}_t \mid \text{Sale}_t\right) = \int p\left(\tilde{\text{Sale}}_t \mid \beta_t, \sigma_t^2\right) p\left(\beta_t, \sigma_t^2 \mid \text{Sale}_t\right) \mathrm{d}\beta_t \mathrm{d}\sigma_t^2 \tag{7}$$

Note that the posterior samples for prediction could sample $\tilde{\text{Sale}}_t$ from the normal distribution based on the posterior distribution of $\widehat{\beta}_t, \widehat{\sigma}_t^2$.

3.2 Evaluation metrics

For prediction, we compute the posterior means of sale $\tilde{\text{Sale}}_{it}$ of every single product in the period t. Mean absolute error (MAE) and root mean squared error (RMSE) are adopted to evaluate the effectiveness of our proposed approach. MAE and RMSE can be calculated as:

$$\text{MAE} = \frac{1}{n} \sum_{i \in n} \left| \text{Sale}_{it} - \tilde{\text{Sale}}_{it} \right| \tag{8}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i \in n} \left(\text{Sale}_{it} - \tilde{\text{Sale}}_{it} \right)^2} \tag{9}$$

where $\tilde{\text{Sale}}_{it}$ and Sale_{it} denote the predicted product sales and sample sales of the i product in e-commerce, respectively, while n is the total number of the products. Obviously, the lower values of MAE and RMSE represent better accuracy.

4. Data description and experiments

4.1 Clickstream data and description

Aiming at forecasting product sales and monitoring product competition using our proposed CBBI model, we acquired a clickstream dataset from one of the top three e-commerce retailers in China. Due to business privacy, the terms of the product and brand were anonymous. The dataset was comprised of each user's recent weblog and was extracted over two weeks from April 1 to April 16, 2016. Each data entry recorded an interaction that occurred when a consumer accessed a product. The frequency of human–computer interactions, consisting of consumers' multiple activities, had more than 13.1 million action points by 78,746 distinct consumers across approximately 200,000 stock-keeping unit products of 399 brands in 8 product categories.

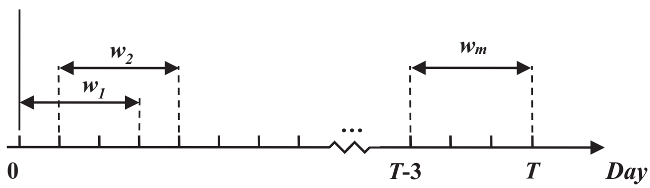


Figure 2.
An illustration of time window $\lambda = 3$ of a sliding window approach

A rich product category with an abundance of products enabled us to generalize our proposed CBBI model. However, in the context of e-commerce, consumers prefer those categories of products that have low outlay and intangible value proposition and are frequently purchased (Phau and Meng Poon, 2000). In other words, online consumer preferences vary across product categories. Specifically, the product category itself determines product market structure and product segments (Erdem, 1996; Netzer *et al.*, 2012). Thus, we implemented the CBBI model disaggregated by product category. In practice, however, it is much more detailed for managers to conduct product management tactics. Table I shows summary statistics of our dataset sorting by product category. Without a loss of generality, this study selected the first category of products as an illustration for analysis while using other categories for robustness testing.

The descriptive statistics for heterogeneous behavioral feedbacks and product sales in the first category of products within half a month are displayed in Table II. Consumer preferences were rather diverse toward a total of 3,158 products, as the standard deviation (SD) (Table II, column 4) of each feedback was rather larger than the corresponding mean (Table II, column 3). Together, they imply that there will likely be segments of the product market in consumers' perceptual space.

4.2 Testing the hypothesis and ascertaining the best time period

A multicollinear test was conducted on the five heterogeneous behavioral feedbacks of each subset and showed that a collinear problem did not exist. To verify the hypothesis that product sales are period-dependent, we conducted the proposed updated BLR and stable BLR to examine the period-dependent effect on predicting sales.

Considering the long-term and short-term memory effect on rational decision-making (Simon, 1979; Yan *et al.*, 2009; Strandburg, 2013), we therefore chronologically divided the clickstream data into subsets with specific time window size λ in the manner of a sliding window (Chu, 1995; Deypir *et al.*, 2012). Next, the initial subset was set as the first-time

Category	No. of brands	No. of products	No. of consumers	No. of records
1	126	3,158	16,351	963,049
2	51	2,496	34,922	3,013,837
3	63	1,995	22,701	1,306,452
4	73	2,057	25,721	15,921,69
5	38	3,077	64,446	5,239,810
6	124	4,304	12,231	928,154
7	89	1,793	4,696	129,194
8	18	628	4,645	27,269

Table I. Summary statistics of the sort-by-product-category dataset

Variable	Product-level definition	Mean	SD	Min	Max
Click	Freq of click	192.8	728.116	0	14,507
Time	Time duration of browsing an item (second)	5,338	20,846.788	0	427,962
Tag	Freq of an item tagged in favorites	0.4971	2.164	0	39
Add	Freq of an item added to the cart	3.459	14.774	0	323
Rem	Freq of an item removed from the cart	1.847	7.544	0	154
Product sale	Proportion of sales of an item in the market (%)	0.0318	0.158	0	2.84

Note(s): In the BLR, we employed the logarithm form of these consumer heterogeneous feedbacks for nondimensionalization, and we added 1 to the logarithmic result of each variable to avoid zero issues

Table II. Statistical description of variables of first category window dataset w_1 for estimating the distributions of parameters $\widehat{\theta}_{w_1} = (\widehat{\beta}_{w_1}, \widehat{\sigma}^2_{w_1})$, which are shown as the yellow blocks in Figure 3. Also, the following subset w_{j+1} with equal size is illustrated as the green blocks in Figure 3, respectively. The prior knowledge in the previous window w_j (yellow blocks in Figure 3) can be utilized to predict the sales in the current window w_{j+1} (green blocks in Figure 3). In the sliding window approach (Chu, 1995; Deypir *et al.*, 2012), the time window moves one block (a day in our study) forward over time, indicating that the following sets dynamically apply the up-to-the-minute data to predict the sales. We considered the size of window λ ranging from 3 to 7 days. Taking the three-day window as an example, the first three days form the initial set w_1, and the consequent set w_{j+1} is applied to predict the sales for the next three days. This form demonstrates the advantage of being able to predict the sales for the following days in a more real-time manner.

4.2.1 Hypothesis testing: period-dependent product sales. Updated and stable BLRs were run to test the hypothesis: Product sales are period-dependent. The results of the stable BLR and the proposed updated BLR with different sizes of time windows λ are depicted in Figure 4. The lower MAE and RMSE values represented a better model fit. In general, the MAE and RMSE values of updated BLR were obviously much lower than those of the stable BLR, shown in Figure 4, respectively, indicating that the updated BLR outperformed the stable BLR no matter if evaluated by MAE or RMSE. The results indicated that estimated knowledge in the previous

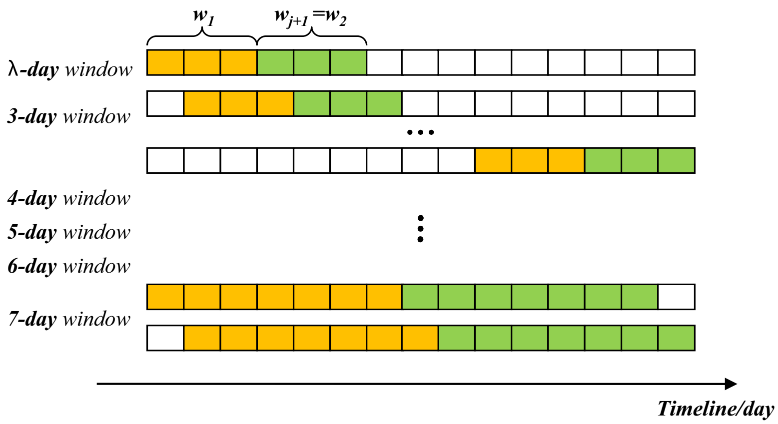


Figure 3.
Sliding time window over time

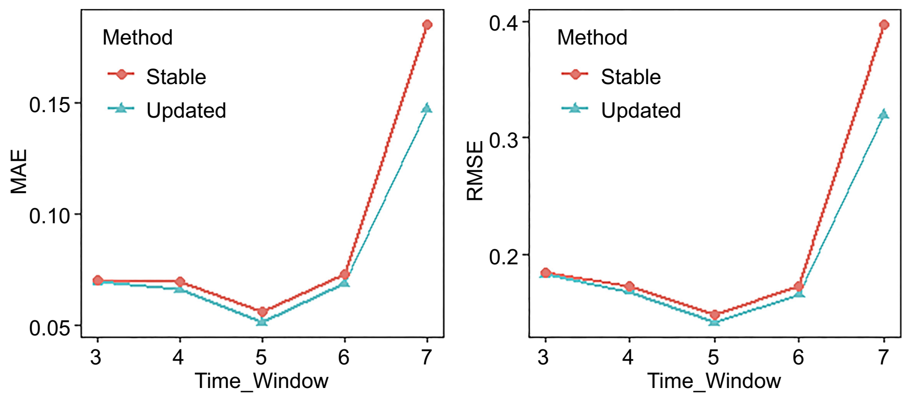


Figure 4.
Performance of different time window settings from stable and updated BLR in MAE (left) and RMSE (right) analyses

period is helpful to infer the sale of products by influencing consumer consequent purchase decisions, thus supporting the hypothesis that product sales are period-dependent.

4.2.2 Ascertaining the best size of a time window. Figure 4 presents the different performances of updated BLR varying from three- to seven-day windows for exogenously ascertaining the best size of time period λ^*. For the MAE value of the updated BLR, as shown in the left of Figure 4, it decreased from the three-day window (0.069) to the five-day window (0.051). However, the curve soared and reached 0.147 at the seven-day window once when the time window exceeded five days. The right of Figure 4 shows that the RMSE value followed a similar pattern. Hence, we found that $\lambda^* = 5$, indicating that the five-day window was the best size of time window for predicting product sales of the sample product category.

Robustness analysis of the other product categories was also conducted, and similar results showed that the product sale is period-dependent, as well as that the best size of the time window was $\lambda^* = 5$, thus representing the most appropriate duration for the memory effect for each product category to forecast and monitor product sales in this focal e-commerce platform.

5. Visualizing competitive product market structure

To gain a better understanding of product marketing, we developed perceptual maps to represent the market structure of the first product category with the final five-day window dataset. We utilized the two conventional unsupervised learning algorithms for visualizing product market structure (Kim *et al.*, 2011; Netzer *et al.*, 2012; Ringel and Skiera, 2016). In other words, we used K-means clustering and the MDS technique to construct the perceptual maps where consumer heterogeneous feedback is elaborated and there is no corresponding dependent variable for the product sale.

Competitive market segmentation and identification of major competitors in a focal segment can gain attention from sellers and manufacturers for the sake of reducing uncertainty in product marketing and management (Grasso, 2018). Thus, we applied a K-means clustering algorithm to address the aforementioned issues. Also, we used the silhouette index (SI), ranging from 0 to 1, as an evaluation index to interpret and verify the consistency of clusters (Savaresi and Boley, 2004). The higher the SI is, the higher cohesion and separation the segments of the market have. The results of SI on the incremental number of submarkets are given in Figure 5. Though the SI of two segments was the highest (SI = 0.469), that of three segments was the optimal point according to the heuristic Elbow principle for practicability (Kodinariya and Makwana, 2013). Robustness analysis for identifying the segments of the market structure related to other product categories was also conducted. Nevertheless, the optimal number of submarkets depended on the product category itself.

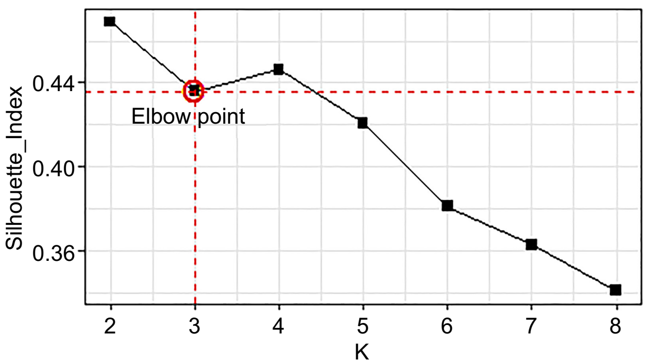


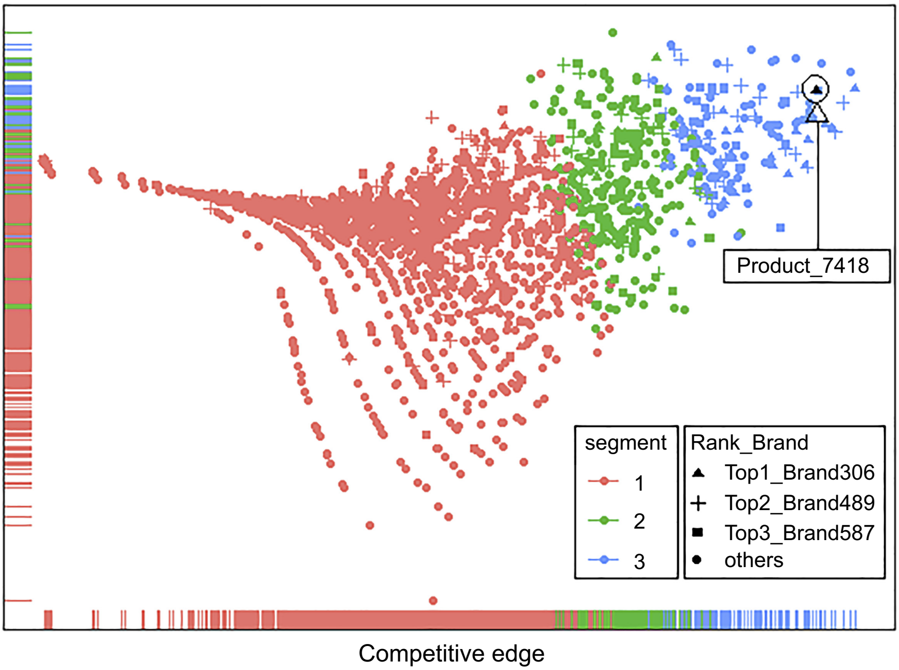
Figure 5.
The number of segments of the market structure

Table III presents the clustering results for the averages of heterogeneous feedback and product sale. The results suggest that product sale is positively associated with the segments, meaning that segment 3 is the most dominant product segment in the focal product market structure while the market share of other segments is rather minor. Furthermore, we mapped the three competitive product segments with heterogeneous consumer feedback, and the map of product market structure we derived is depicted in Figure 6. Each of the scatter points in the figure is a specific product in the first product category. We added an overlay (scatter color, i.e. red, green, and blue, respectively) that shows market segments derived through the K-means clustering algorithm, and the scatters with the shapes of a triangle, cross, and square represent the products from the top-three brands, respectively. Although the two dimensions cannot have a single and global meaning across the clusters, it was surprising to find that the closer a focal product point was to the positive of the horizontal axis, the stronger its competitive degree was in consumers' perceptual space. Irrespective of product sale, our

Segment	log_Click	log_Time	log_Tag	log_Add	log_Rem	Product sale
Segment1	2.251	4.513	0.020	0.113	0.087	0.003
Segment2	4.928	8.228	0.237	1.167	0.712	0.062
Segment3	6.572	9.920	0.903	2.742	2.083	0.543

Note(s): Consumer heterogeneous feedback was considered as input in the unsupervised learning algorithm K-means clustering while product sale was not a consideration

Table III.
Means of variables of segments



Note: Consumer heterogeneous feedback was as input in the unsupervised learning algorithm MDS while product sale was not a consideration

Figure 6.
Product map of consumer–product interaction in consumer perceptual space

visualization results indicate that the horizontal axis could be represented as a signal of the ascending trend of product sale in Figure 6.

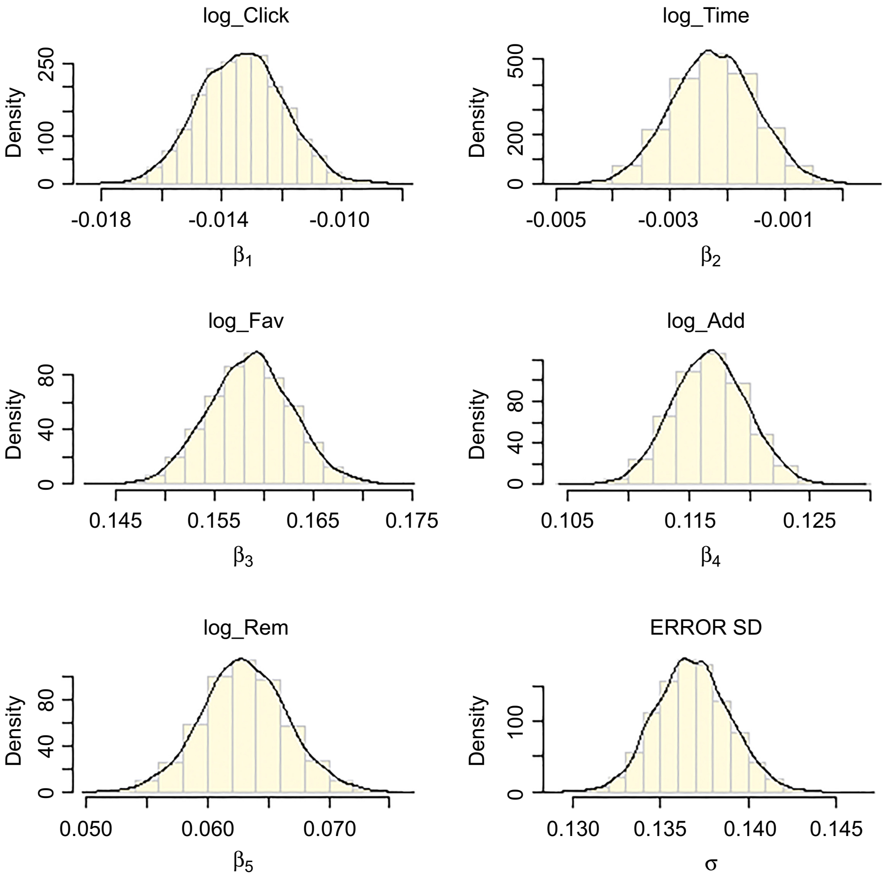
For product-level market structure, product proximity is the basic feature of identifying the market clusters shown in the product map. Smaller Euclidean distances represent higher product similarity. We see that three distinguishable market segments are built with regard to products in close proximity. Visually drawn from Figure 6, the two areas shaped by the red scatters for segment 1 (consisting of 1,750 unique products) and those shaped by the green for segment 2 (264 correspondingly) are much larger than the blue area for segment 3 (only 145 products). Statistically, 6.7 percent of the products in segment 3 account for 78.68 percent of product sales in the whole market, while most products present very little competition, which is close to the 80/20 rule, verifying the known Pareto principle in e-commerce (Fujiwara *et al.*, 2004; Zuo *et al.*, 2019).

On the other hand, we found that there was an overlapped scope between the adjacent segments shown by the color-overlapped rug plots at the left and bottom of Figure 6. Even though the horizontal axis of the MDS as a signal of the ascending trend of product sales the best-seller product_7418 had was the most competitive (2.36 percent of product sales in the product market), but was not at the farthest position to the right on the plot. This finding illustrates the heterogeneous preferences across consumers, suggesting that consumer heterogeneity could be further explored for mining business intelligence (Ding *et al.*, 2015), especially for measuring market structure more precisely. The segments indicated that different levels of competition are driven by consumer perceptual preferences, providing a more intuitive insight to help in effectively scouting for deadly competitors in different product submarkets.

In respect to the brand-level market structure, we took the three top brands as examples for particular managerial implications. We found that products of the three top brands were mainly distributed in segments 2 and 3, shown as the scatters with the shape of a triangle, cross, and square in Figure 6, where the brands no.306 and no.489 had dominant positions in the entire product market and whose overall product sales were 19.62 percent and 17.69 percent, respectively. But, the two brands had 19.41 percent and 17.06 percent, respectively, in segment 3, both accounting for over 95 percent of product sales by themselves (19.41/19.62 percent and 17.06/17.69 percent, respectively) in that product market. For the managers, integrating lean and sustainable supply chain management could benefit from appropriate product cutdown (Zhu *et al.*, 2018) to achieve cost-effectiveness. Nonetheless, adopting a product-diversified managerial strategy to hedge competitive risks (Peterson *et al.*, 1997) is equally as vital. The top-three brands, no.587, had 9.82 percent of the product sales in segment 3 and overall, 12.99 percent of product sales in the entire market. However, notably, brand no.587 was still the dominant brand in segments 1 and 2. In other words, the brand owner might adopt a product-diversified marketing strategy to accelerate sales for products belonging to brand no.587.

6. Parameter estimation by updated BLR

For a clear expression, a multicollinear test was again conducted in the final five-day window. Variance inflation factor (VIF) analysis showed that no collinear problem existed among all behavioral feedbacks, specifically, click ($4.668 < 10$), time-of-browsing ($3.150 < 10$), tag-into-favorite ($1.799 < 10$, add-into-cart ($5.166 < 10$), and remove-from-cart ($4.195 < 10$). After ascertaining the best period, we conducted our proposed updated BLR for the final five-day-window dataset to again present the results of parameter estimation. More specifically, we ran a Bayesian inference algorithm for 100,000 iterations in the *R* programming language (the *R*-3.4.1 version). The first 95,000 iterations served as the "burn-in" period to ensure convergence, and the last 5,000 iterations served as the input for our parameter inference. Compared with the OLS method of point estimation, the updated BLR outputs the posterior distribution of parameters $\widehat{\theta}_t = (\widehat{\beta}_t, \widehat{\sigma}_t^2)$, as shown in Figure 7.



Note: Here, log_Click, log_Time, log_Tag, log_Add, and log_Rem represent the logarithm form of heterogeneous feedback, respectively, denoted in Table 2

Figure 7.
Posterior distribution of parameters $\widehat{\theta}_t = (\widehat{\beta}_t, \widehat{\sigma}_t^2)$

Based on what is shown in Figure 7, we can deduce that it is intuitive that each heterogeneous feedback is significantly associated with product sale, since zeros do not lie in their 95 percent posterior probability interval, which is in line with the rule of the BLR (Rossi and Allenby, 2003; Albert, 2009). Interestingly, taking log_Click as an example, we commonly think that the more frequently a consumer clicks a focal product, the higher is the likelihood that she will buy it. However, the posterior distribution of parameter β_1 was stabilized from –0.0182 to –0.0084, implying that consumers might reveal dislikes toward products in the course of online shopping, turning into negative outcomes on product sale. The SD $\widehat{\sigma}$ of sale of products was large and stable at 0.137 as compared with the average 0.0318 product sale, representing much larger differences in degree of competition across products. The result was consistent with the findings for segments of the product market structure shown in Section 5.

We exemplify the posterior distribution of $\widehat{\theta}_t = (\widehat{\beta}_t, \widehat{\sigma}_t^2)$ in Table IV. We can conclude from the data displayed in Table IV that click and time-of-browsing have a significant negative impact on product sale ($\widehat{\beta}_1 = -0.013$, $\widehat{\beta}_2 = -0.002$). The findings show that a high level of consumer engagement in products is not necessary to complete transactions, which might be owing to the fluctuant human interest affected by the occurrence of events (Han *et al.*, 2008; Ding *et al.*, 2015). For instance, consumers have hedonic shopping motivations impacting their online impulse buying behavior during a comparatively short shopping trip (Ozen and Engizek, 2014). Nonetheless, when consumers shift to hedonic motivation, they will click more and browse more products but not necessarily make a transaction. Thus, consumer heterogeneity should be considered in the online environment more so than in its brick-and-mortar counterpart (Fu *et al.*, 2013), which might lead to interpreting the negative effects of click and time-of-browsing and considering under what boundary conditions such feedback will change the valence of effects.

However, products tagged-into-favorite ($\widehat{\beta}_3 = -0.159$) or added-into-cart ($\widehat{\beta}_4 = -0.117$) are more likely to have a higher degree of competitiveness in market. Intuitively, the sale of products depends on the level of consumer preference for the products. This finding is in accordance with the case of tagging behavior in Taobao.com (Ou and Chan, 2014). The consumer tags the item to favorites, which represents a positive signal on the probability to purchase the item and could be exploited to shape the competitive edge of a seller (Olbrich and Holsing, 2011; Ou and Chan, 2014).

Besides consumer preferences, consumer deletion behavior needs to be considered when analyzing product markets (Day *et al.*, 1979). However, the consumer remove-from-cart function is positively and significantly associated with product sale ($\widehat{\beta}_5 = -0.063$), which implies that consumers continue to have purchase intention of a choice alternative for one that was removed. In fact, many consumers use their online shopping cart as a tool to help gather information. Even though tag and add-into-cart serve as indicators of desires of consumers to buy, the feedback remove-from-cart conveys information about the consumer consideration set to sellers or product managers (Kukar-Kinney and Close, 2010). Such information may delay consumers' purchase decisions but still give an implication that the product removed is still in the consumer consideration set.

7. Discussion and implications

7.1 Key findings

The present study proposes a CBBI model to highlight the role of consumer heterogeneous feedback and the memory effect of rational decision-making on visualizing product market structure and predicting dynamic product competition in terms of product sales. The results of the study provided significant findings.

Particularly in the prediction part of the CBBI model, the proposed updated BLR incorporating prior historical knowledge outperformed a stable BLR in terms of MAE and RMSE analyses, suggesting that product sale is period-dependent due to the consumer memory effect. Furthermore, through the sliding window approach (Deypir *et al.*, 2012), the evaluation of MAE showed that the performance of different sizes of the

	Mean	Estimated coefficient 5%	50%	95%	Standard error
Intercept	*0.013*	0.009	0.013	0.018	0.0028
log_Click	*−0.013*	−0.016	−0.013	−0.011	0.0014
log_Time	*−0.002*	−0.003	−0.002	−0.001	0.0007
log_Tag	*0.159*	0.152	0.159	0.166	0.0042
log_Add	*0.117*	0.112	0.117	0.122	0.0031
log_Rem	*0.063*	0.057	0.063	0.069	0.0035
Error sd	0.137	0.134	0.137	0.140	/

Note(s): Estimates in bold indicate that zero does not lie in the 95 percent posterior probability interval

Table IV. Posterior distribution of parameter estimates time window setting decreased from a three-day window (0.069) to a five-day window (0.051), and the curve soared and reached 0.147 at a seven-day window. RMSE presented a similar pattern. Robustness analysis of other product categories also ascertained that the size of five-day time window ($\lambda^* = 5$) was the most appropriate duration for the memory effect for each product category to forecast product sales in the focal e-commerce platform.

Our findings from the description part of CBBI model utilizing the K-means clustering and MDS technique with consumer heterogeneous feedback showed that three market segments are optimal for identifying the product market structure. As shown in Figure 6, it was surprising to find that the closer a focal product point was to the positive of the horizontal axis, the stronger its competitive degree was in consumers' perceptual space. In particular, segment 1 was the least competitive edge segment, and segment 3 was the most dominant market segment. That proved to be intuitive for scouting for major product or brand competitors in each segment. Overall, the description part of CBBI initially indicated that consumer heterogeneous feedback is associated with product sale.

Moreover, for product-level market structure, 6.7 percent of products in segment 3 accounted for 78.68 percent of product sales in the entire market while most products had very little competitive advantage, which was close to the 80/20 rule, verifying the known Pareto principle in e-commerce. As to the brand-level market structure, the two top brands were mainly distributed in the most competitive segment 3, revealing a lean and sustainable supply chain management. At the same time, the top-three brands were scattered widely in each segment, indicating a product-diversified managerial strategy.

To further scrutinize the effects of consumer heterogeneous feedback on product sales, our research employed the updated BLR to demonstrate that click and time-of-browsing have significant negative impacts on product sales ($\widehat{\beta}_1 = 0.013$, $\widehat{\beta}_2 = 0.002$). However, products tagged-into-favorite ($\widehat{\beta}_3 = 0.159$) or added-into-cart ($\widehat{\beta}_4 = 0.117$) were found to be more likely to have a higher degree of global competitiveness in the market. Interestingly, consumer remove-from-cart was positively and significantly associated with product sales ($\widehat{\beta}_5 = 0.063$). The implications of these key findings are discussed in the following section.

7.2 Theoretical implications

This study is among one of the first that focuses on consumer heterogeneous feedback in product market structure. Most all of the previous studies have focused on homogeneous feedback to monitor the competitive edge of products in single search data (Bao *et al.*, 2008; Lee and Bradlow, 2011; Netzer *et al.*, 2012; Ringel and Skiera, 2016; Wei *et al.*, 2016), This study highlights the competitive intelligence of consumer behavior, that is, consumers' click, time-of-browsing, tag-into-favorite, add-into-cart, remove-from-cart, in determining a product's competitive edge in terms of product-level sales. By extending the literature of product market structure from the perspective of consumer feedback data (Kim *et al.*, 2011; Netzer *et al.*, 2012; Ringel and Skiera, 2016), this study proposes a CBBI model to visualize product market structure and predict dynamic product competition using clickstream data. Thus, a number of theoretical insights were provided by this study.

Firstly, this study contributes to the existing studies on the consumer memory effect and product competition. In terms of the discipline of operation management, Azoury and Miyaoka (2009) found that the state-dependent policy is optimal for forecasting product demand and inventory production. In marketing science, previous studies (Yan *et al.*, 2009; Strandburg, 2013) were largely concerned about the consumer memory effect on marketing strategies. Instead of merely focusing on the consumer memory effect, this study utilized prior knowledge to infer consumer consequent purchase decisions resulting in the degree of product-level sale, namely, the competitive edge of individual products. Robustness analysis of each product category suggested that product sale is period-dependent in e-commerce. On the other hand, the size of the time window $\lambda^* = 5$ was found to be the most appropriate duration for the memory effect to forecast product sales from the perspective of the sliding window of time series. This study also suggests a new integrated theoretical lens for considering the consumer memory effect based on rational decision-making (Simon, 1979) to examine the product competition.

From the theoretical perspective, this study extends the research on homogeneous feedback for visualization of product market structure (Kim *et al.*, 2011; Netzer *et al.*, 2012; Ringel and Skiera, 2016). Akin to these studies (Netzer *et al.*, 2012; Ringel and Skiera, 2016), without considering product sales, the conventional K-means clustering and MDS incorporate consumer heterogeneous feedback and indicate that those factors could reveal the competitive degree of products by using the horizontal axis in the map of consumers' perceptual space. Thus, this study provides a novel perspective that incorporates consumer heterogeneous feedback to identify market structure.

In terms of the different effects of consumer heterogeneous feedback, this study also extends the stream of research (Bao *et al.*, 2008; Lee and Bradlow, 2011; Wei *et al.*, 2016) on homogeneous feedback to predict the competitive degree of products. It also extends and validates the related research (Olbrich and Holsing, 2011; Ding *et al.*, 2015), indicating that the more consumers click a product, the deeper the product is visited and the lower the probability of purchase. The time spent on a product also represents the same negative impact on purchase probability. The feedback of tag-into-favorite is positively associated with product sales, as shown by the studies (Olbrich and Holsing, 2011; Ou and Chan, 2014). Thus, tags can be a useful social signal for making the shopping journey more effective through helping consumers to access what they demand. Similarly, the shopping cart is identified as the consideration set for consumers: when consumers add a product into shopping cart, this indicates their intention to purchase such a product. Shopping cart abandonment is highest among consumers who have a hedonic motivation or are planning a future purchase and simply gathering information (Day *et al.*, 1979; Kukar-Kinney and Close, 2010). Notably, the products that have been abandoned still hold positive significance for further transactions. Hence, this study provides researchers with a springboard to further integrate consumer preferences, product deletion, and addition behavior to investigate product competition.

7.3 Managerial implications

The results also have valuable practical implications. First, they suggest that the memory effect could be valuable to sellers for sales prediction in the platform of e-commerce. As the number of products grows enormous, the consequent abundance of product information results in dynamic and competitive circumstances that require sellers to adjust their product management strategies in a timely fashion in accordance with intensive competition in market structure (Grasso, 2018). The study shows that incorporating the memory effect can potentially improve product sale prediction accuracy. At the same time, the updated BLR allows sellers to conduct a sliding window approach to predict the sales in a time period that their personalization demands to prepare a better strategy earlier to avoid the worst conditions and manage the risk in the future.

Second, the visualization part of CBBI provides sellers and managers with an intuitive way to find the position where their products or brands stand. As informed by the perceptual product map, its horizontal axis, which serves as a signal of the ascending trend of product sales, enables sellers to position their products and quickly scout for their major competitors in consumer perceptual space in a straightforward manner. Furthermore, the three distinguishing market segments displayed in different competitive degrees could help sellers choose their products based on what is popular in a specific market segment and what is the major product substitution.

Despite the existence of the Pareto principle in the market structure, we found that the corresponding lower number of products contributed to the major sales in the entire market. Products belonging to the top-three brands, no.587, also were dominantly distributed in each segment, suggesting that adopting a product-diversified managerial strategy to hedge competitive risks (Peterson *et al.*, 1997) is vital for sellers. In such a case, sellers and product managers might adjust their marketing strategies to satisfy consumers' varied demands. On the other hand, products belonging to the top-two brands, no.306 and no.489, merely were in the most competitive edge segment. Thus, for their sellers and managers, integrating lean and sustainable supply chain management (Zhu *et al.*, 2018) could be beneficial for achieving cost-effectiveness. Specifically, it is beneficial to cut down product supply in the other two segments and make accurate marketing strategies to boost product sales. However, since the number of market segments depends on the product category, sellers and managers should be cautious about generalizing the two strategies from one product category to another.

Third, the detailed distribution of parameters $\widehat{\theta}_t = (\widehat{\beta}_t, \widehat{\sigma}_t^2)$ from the updated BLR could be highly informative for product managers and manufacturers when conducting further simulations of product sales by controlling some variables. For instance, sellers and product managers could control some behavioral variables constantly and simulate the sale of a focal product. This would be a novel perspective when scouting for new products according to popularity across consumers in order to resist competitive risk. The current e-commerce has heavily applied the collaborative filter algorithm to achieve personalization marketing, through recommending consumers' popular products based on those consumers who have similar transaction histories. However, our findings suggest that personalization marketing tactics could incorporate not only consumer similarity but also a social commerce mechanism and consideration set within shopping cart. Tagging a product into favorites has a positive impact, and it is a social signal to attract potential buyers to buy such a product. A better webpage design for personalization marketing would be to include some related high-frequency tagged-into-favorite products to draw more attention and make the search process more effective, as well as to help consumers to find what they indeed need. Our results also demonstrate that both the products added-into-cart and those removed-from-cart have a positive influence on product sales. Thus, sellers and retailers need to pay attention to the consideration set in the shopping cart. Akin to the implication of tagging behavior, when consumers search products, the page of the searching results could display the related products that have been added to cart, as well as highlight their characteristics. Consumers tend to use their cart as a consideration set to store their desired products, to track prices, or for other purposes, which might cause a delay in making a purchase (Kukar-Kinney and Close, 2010). Thus, abandoned carts are still valuable for both consumers and sellers. The specific products removed-from-cart also provide sellers and managers an opportunity to understand information about consumers' consideration sets. On the one hand, sellers and managers could still recommend such products to consumers while their prices change to gain profit. On the other hand, it is likely that this would provide alternatives or complementary products with which to target consumers.

8. Limitations and future research directions

This research is subject to several limitations. First, consumer heterogeneity of clickstream data is not fully considered in this study. This might have had some influence on why consumer click and time of-browsing have negative impacts on product sales and under what boundary conditions these two types of feedback could have opposite effects. These negative impacts may indicate that, when consumers shift to hedonic motivation, they will click products more and spend more time without making a transaction. Further research is encouraged to take consumer heterogeneity into consideration. Second, seasonality could influence product sales of different product categories, especially clothing products. However, seasonality was not considered as an input variable to predict the product sales in the model. It would also be relevant to further explore the effect of seasonality to measure market structure more precisely. More data on both of these factors could enhance the empirical findings of the current study.

Highlights

- We propose a consumer-behavior-based intelligence (CBBI) model to identify market structure so as to monitor product competition, using real-world clickstream data from top-two e-commerce retailers in China.
- By taking users' online behaviors into account, the research develops an updated Bayesian linear regression (BLR) that outperforms the conventional BLR in terms of mean absolute error (MAE) and root mean squared error (RMSE) analyses, thus supporting the hypothesis that online product sales are period-dependent owing to the consumer memory effect in e-commerce.
- Based on consumers' heterogeneous behavioral feedback, the current research visualizes three different segments of the competitive market structure in a perceptual map, its horizontal axis shown as a signal of the ascending trend of product sales.
- The estimated results of the proposed updated BLR imply that consumers' online behavioral feedback of tag-into-favorite, add-into-cart, and remove-from-cart have positive impacts on product-level sales in e-commerce, while other types of user shopping status, such as click and time-of-browsing, have negative effects on sales.
- The optimal number of market segments depends on the product category itself.

References

Albert, J. (2009), "Regression models", *Bayesian Computation with R*, Springer, New York, NY, pp. 205-234.

Alfian, G. (2019), "Customer behavior analysis using real-time data processing", *Asia Pacific Journal of Marketing and Logistics*, Vol. 31 No.1, pp. 265-290.

Ansari, A. and Mela, C.F. (2003), "E-customization", *Journal of Marketing Research*, Vol. 40 No. 2, pp. 131-145.

Azoury, K.S. and Miyaoka, J. (2009), "Optimal policies and approximations for a bayesian linear regression inventory model", *Management Science*, Vol. 55 No. 5, pp. 813-826.

Bao, S., Li, R., Yu, Y. and Cao, Y. (2008), "Competitor mining with the web", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 20 No. 10, pp. 1297-1310.

Bonfrer, A. and Drèze, X. (2009), "Real-time evaluation of e-mail campaign performance", *Marketing Science*, Vol. 28 No. 2, pp. 251-263.

Bruwer, J., Roediger, B. and Herbst, F. (2017), "Domain-specific market segmentation:a wine-related lifestyle (WRL) approach", *Asia Pacific Journal of Marketing and Logistics*, Vol. 29 No. 1, pp. 4-26.

Bucklin, R.E. and Sismeiro, C. (2009), "Click here for internet insight: advances in clickstream data analysis in marketing", *Journal of Interactive Marketing*, Vol. 23 No. 1, pp. 35-48.

Chu, C.-S.J. (1995), "Time series segmentation: a sliding window approach", *Information Sciences*, Vol. 85 Nos 1-3, pp. 147-173.

Day, G.S., Shocker, A.D. and Srivastava, R.K. (1979), "Customer-oriented approaches to identifying product-markets", *Journal of Marketing*, Vol. 43 No. 4, pp. 8-19.

DeSarbo, W.S. and Grewal, R. (2007), "An alternative efficient representation of demand-based competitive asymmetry", *Strategic Management Journal*, Vol. 28 No. 7, pp. 755-766.

DeSarbo, W.S., Grewal, R. and Wind, J. (2006), "Who competes with whom? A demand-based perspective for identifying and representing asymmetric competition", *Strategic Management Journal*, Vol. 27 No. 2, pp. 101-129.

Deypir, M., Sadreddini, M.H. and Hashemi, S. (2012), "Towards a variable size sliding window model for frequent itemset mining over data streams", *Computers and Industrial Engineering*, Vol. 63 No. 1, pp. 161-172.

Ding, A.W., Li, S. and Chatterjee, P. (2015), "Learning user real-time intent for optimal dynamic web page transformation", *Information Systems Research*, Vol. 26 No. 2, pp. 339-359.

Ehrenberg, A.S., Uncles, M.D. and Goodhardt, G.J. (2004), "Understanding brand performance measures: using dirichlet benchmarks", *Journal of Business Research,* Vol. 57 No. 12, pp. 1307-1325.

Elrod, T., Russell, G.J., Shocker, A.D., Andrews, R.L., Bacon, L., Bayus, B.L., Carroll, J.D., Johnson, R.M., Kamakura, W.A. and Lenk, P. (2002), "Inferring market structure from customer response to competing and complementary products", *Marketing Letters*, Vol. 13 No. 3, pp. 221-232.

Erdem, T. (1996), "A dynamic analysis of market structure based on panel data", *Marketing Science*, Vol. 15 No. 4, pp. 359-378.

France, S.L. and Ghose, S. (2016), "An analysis and visualization methodology for identifying and testing market structure", *Marketing Science*, Vol. 35 No. 1, pp. 182-197.

Fu, Y., Lai, K.K. and Liang, L. (2013), "Bricks or clicks", *Asia Pacific Journal of Marketing and Logistics*, Vol. 25. No. 4, pp. 695-714.

Fujiwara, Y., Di Guilmi, C., Aoyama, H., Gallegati, M. and Souma, W. (2004), "Do pareto–zipf and gibrat laws hold true? An analysis with European firms", *Physica A: Statistical Mechanics and Its Applications*, Vol. 335 Nos 1-2, pp. 197-216.

Glushko, R.J., Maglio, P.P., Matlock, T. and Barsalou, L.W. (2008), "Categorization in the wild", *Trends in Cognitive Sciences*, Vol. 12 No.4, pp. 129-135.

Grasso, C. (2018), "The state of the amazon marketplace 2018", available at: https://fv.feedvisor.com/rs/656-BMZ-780/images/Feedvisor_The-State-of-the-Amazon-Marketplace.pdf (accessed 26 December 2018).

Han, X.-P., Zhou, T. and Wang, B.-H. (2008), "Modeling human dynamics with adaptive interest", *New Journal of Physics*, Vol. 10 No. 7, 073010.

Ijaz, M. and Rhee, J. (2018), "Constituents and consequences of online-shopping in sustainable e-business: an experimental study of online-shopping malls", *Sustainability*, Vol. 10 No. 10, p. 3756.

Ijaz, M., Tao, W., Rhee, J., Kang, Y.-S. and Alfian, G. (2016), "Efficient digital signage-based online store layout: an experimental study", *Sustainability*, Vol. 8 No. 6, pp. 511.

Ijaz, M.F., Rhee, J., Lee, Y.-H. and Alfian, G. (2014), "Efficient digital signage layout as a replacement to virtual store layout", *International Journal of Information and Electronics Engineering*, Vol. 4 No. 4, p. 312.

Kim, J.B., Albuquerque, P. and Bronnenberg, B.J. (2011), "Mapping online consumer search", *Journal of Marketing Research*, Vol. 48 No. 1, pp. 13-27.

Kodinariya, T.M. and Makwana, P.R. (2013), "Review on determining number of cluster in K-means clustering", *International Journal*, Vol. 1 No. 6, pp. 90-95.

Kukar-Kinney, M. and Close, A.G. (2010), “The determinants of consumers’ online shopping cart abandonment”, *Journal of the Academy of Marketing Science,* Vol. 38 No. 2, pp. 240-250.

Lee, T.Y. and Bradlow, E.T. (2011), “Automated marketing research using online customer reviews”, *Journal of Marketing Research*, Vol. 48 No. 5, pp. 881-894.

Leszczyc, P.T.P., Sinha, A. and Timmermans, H.J. (2000), “Consumer store choice dynamics: an analysis of the competitive market structure for grocery stores”, *Journal of Retailing*, Vol. 76 No. 3, pp. 323-345.

MacInnis, D.J. (2011), “A framework for conceptual contributions in marketing”, *Journal of Marketing*, Vol. 75 No. 4, pp. 136-154.

Moe, W.W. and Fader, P.S. (2004), “Dynamic conversion behavior at e-commerce sites”, *Management Science,* Vol. 50 No. 3, pp. 326-335.

Montgomery, A.L., Li, S., Srinivasan, K. and Liechty, J.C. (2004), “Modeling online browsing and path analysis using clickstream data”, *Marketing Science*, Vol. 23 No. 4, pp. 579-595.

Netzer, O., Feldman, R., Goldenberg, J. and Fresko, M. (2012), "Mine your own business: market-structure surveillance through text mining", *Marketing Science*, Vol. 31 No. 3, pp. 521-543.

Nishimura, N., Sukegawa, N., Takano, Y. and Iwanaga, J. (2018), “A latent-class model for estimating product-choice probabilities from clickstream data”, *Information Sciences*, Vol. 429, pp. 406-420.

Olbrich, R. and Holsing, C. (2011), “Modeling consumer purchasing behavior in social shopping communities with clickstream data”, *International Journal of Electronic Commerce*, Vol. 16 No. 2, pp. 15-40.

Ou, C.X. and Chan, K.C. (2014), “Developing a competitive edge in electronic markets via institutional and social based quality signaling mechanisms”, *Information and Management,* Vol. 51 No. 5, pp. 532-540.

Ozen, H. and Engizek, N. (2014), “Shopping online without thinking: being emotional or rational?”, *Asia Pacific Journal of Marketing and Logistics,* Vol. 26 No. 1, pp. 78-93.

Peterson, R.A., Balasubramanian, S. and Bronnenberg, B.J. (1997), “Exploring the implications of the internet for consumer marketing”, *Journal of the Academy of Marketing Science*, Vol. 25 No. 4, p. 329.

Phau, I. and Lum, L. (2000), “Effects of “physical attractiveness” in the evaluation of print advertisements”, *Asia Pacific Journal of Marketing and Logistics*, Vol. 12 No. 4, pp. 41-59.

Phau, I. and Meng Poon, S. (2000), “Factors influencing the types of products and services purchased over the internet”, *Internet Research*, Vol. 10 No. 2, pp. 102-113.

Porter, M. (1979), “How competitive forces shape strategy”, *Harvard Business Review*, Vol. 57 No. 2, pp. 137-145.

Raphaeli, O., Goldstein, A. and Fink, L. (2017), “Analyzing online consumer behavior in mobile and PC devices: a novel web usage mining approach”, *Electronic Commerce Research and Applications*, Vol. 26, pp. 1-12.

Ringel, D.M. and Skiera, B. (2016), “Visualizing asymmetric competition among more than 1,000 products using big search data”, *Marketing Science*, Vol. 35 No. 3, pp. 511-534.

Rossi, P.E. and Allenby, G.M. (2003), “Bayesian statistics and marketing”, *Marketing Science*, Vol. 22 No. 3, pp. 304-328.

Rutz, O.J. and Bucklin, R.E. (2012), “Does banner advertising affect browsing for brands? clickstream choice model says yes, for some”, *Quantitative Marketing and Economics,* Vol. 10 No. 2, pp. 231-257.

Savaresi, S.M. and Boley, D.L. (2004), “A comparative analysis on the bisecting K-means and the PDDP clustering algorithms”, *Intelligent Data Analysis*, Vol. 8 No. 4, pp. 345-362.

Simon, H.A. (1979), “Rational decision making in business organizations”, *The American Economic Review*, Vol. 69 No. 4, pp. 493-513.

Sismeiro, C. and Bucklin, R.E. (2004), "Modeling purchase behavior at an e-commerce web site: a task-completion approach", *Journal of Marketing Research*, Vol. 41 No. 3, pp. 306-323.

Strandburg, K.J. (2013), "Free fall: the online market's consumer preference disconnect", *University of Chicago Legal Forum*, Vol. 2013 No. 1, p. 5.

Su, Q. and Chen, L. (2015), "A method for discovering clusters of e-commerce interest patterns using click-stream data", *Electronic Commerce Research and Applications*, Vol. 14 No. 1, pp. 1-13.

Su, Q. and Liu, W. (2017), "Market segmentation and pricing strategies based on product platform", *Asia Pacific Journal of Marketing and Logistics*, Vol. 29 No. 1, pp. 183-199.

Sultan, P., Wong, H.Y. and Sigala, M. (2018), "Segmenting the Australian organic food consumer market", *Asia Pacific Journal of Marketing and Logistics*, Vol. 30 No. 1, pp. 163-181.

Urban, G.L., Johnson, P.L. and Hauser, J.R. (1984), "Testing competitive market structures", *Marketing Science*, Vol. 3 No. 2, pp. 83-112.

Wei, Q., Qiao, D., Zhang, J., Chen, G. and Guo, X. (2016), "A novel bipartite graph based competitiveness degree analysis from query logs", *ACM Transactions on Knowledge Discovery from Data (TKDD)*, Vol. 11 No. 2, p. 21.

Wen, K.-W. and Peng, K.-F. (2002), "Market segmentation via structured click stream analysis", *Industrial Management and Data Systems*, Vol. 102 No. 9, pp. 493-502.

Wu, S.I. (2002), "Internet marketing involvement and consumer behavior", *Asia Pacific Journal of Marketing and Logistics*, Vol. 14 No. 4, pp. 36-53.

Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y. and Chen, Z. (2009), "How much can behavioral targeting help online advertising?", *Proceedings of the 18th International Conference on World Wide Web in Madrid*, Spain, 2007, ACM, pp. 261-270.

Zellner, A. (1976), "Bayesian and non-bayesian analysis of the regression model with multivariate student-t error terms", *Journal of the American Statistical Association,* Vol. 71 No. 354, pp. 400-405.

Zuo, M., Liu, H., Zhu, H. and Gao, H. (2019), "Dynamic property of consumer-based brand competitiveness (CBBC) in human interaction behavior", *Industrial Management and Data Systems*, Vol. 119 No. 6, pp. 1223-1241.

Zhu, Q., Shah, P. and Sarkis, J. (2018), "Addition by subtraction: integrating product deletion with lean and sustainable supply chain management", *International Journal of Production Economics,* Vol. 205, pp. 201-214.

About the authors

Mingjun Zhan is a PhD candidate at the School of Management, Guangdong University of Technology, China, majored in Management Science and Engineering. His research interests include consumer behavior, data mining, and business intelligence. His works have been published by journals and conference such as Data Analysis and Knowledge Discovery, International Conference on Electronic Commerce (ICEC), Pacific Asia Conference on Information Systems (PACIS), and Americas Conference on Information systems (AMCIS).

Hongming Gao* is a PhD candidate at School of Management, Guangdong University of Technology, China, majored in Management Science and Engineering. His research interests include consumer behavior, data mining, and electronic commerce. He was a visiting scholar in Tilburg University, Netherlands and a current student research assistant in the University of Hong Kong, China. His works have been published in such journals as Industrial Management and Data Systems, Data Analysis and Knowledge Discovery, Chinese Journal of Management Science, Journal of Guangdong University of Technology, among others. His works have appeared in conferences such as Americas Conference on Information systems (AMCIS), Pacific Asia Conference on Information Systems (PACIS), and International Conference on Electronic Commerce (ICEC). His ORCID is https://orcid.org/0000-0001-6273-6695. Hongming Gao is the corresponding author and can be contacted at: hm.gao.normal@gmail.com

Hongwei Liu is professor of information management at the Department of Management Science and Engineering, Guangdong University of Technology, China. His research interests include business

intelligence, systems design, and privacy issues of data mining. His publications have appeared in such journals as Information and Management, Information Systems, among others. Hongwei Liu was the dean at the School of Management and in charge of the academic research. He also holds several projects of Natural Science Foundation of China (NSFC).

Yidan Peng is a master's student at the School of Management, Guangdong University of Technology, China, majored in Business Administration. Her research interests include consumer behavior and electronic commerce.

Dan Lu is a lecturer at the School of Management, Guangdong University of Technology, China; She is also a PhD candidate majored in Management Science and Engineering. Her research interests include consumer behavior, electronic commerce.

Hui Zhu is lecturer at the Electronic Commerce Department, Guangzhou University, China. Her research interests include knowledge management, electronic commerce, data mining, systems design, and privacy issues of data mining. Her works have been published in journals such as Information and Management, Information Systems, International Journal of Information Processing and Management, and Journal of Network and Information Security.