Electronic Commerce Research
https://doi.org/10.1007/s10660-019-09368-1

Analysis of launch strategy in cross-border e-Commerce market via topic modeling of consumer reviews

Feifei Wang[1,2,5] · **Yang Yang**[3] · **Geoffrey K. F. Tso**[4] · **Yang Li**[1,2,5]

Abstract

Spurred by the policy of China's Belt and Road Initiative, Chinese e-Commerce companies have found great opportunities in selling goods overseas. The cross-border e-Commerce shares similarities of launch and marketing strategies with domestic e-Commerce, but also has substantial differences. How to make strategic adjustments to better adapt to the overseas market is of great concern to cross-border e-Commerce companies. Analyzing behaviors of overseas consumers could offer an effective way to address this issue and has attracted great interest of researchers. Consumer comments, cheap and abundant by its nature, provides an easy access for analysis of consumer behaviors. In this paper, we focus on consumer reviews of a specific product, the cellphones, and apply topic modeling techniques to investigate the differences between behaviors of domestic and overseas consumers. We find that consumers from domestic and overseas focus on different aspects of product. In addition, the foreign consumers care more about product quality and tend to make description of technique details. On the contrary, domestic buyers pay more attention on consumer services and intend to comment in generalities. All these findings could help e-Commerce companies design better launch strategies in cross-border e-Commerce market.

Keywords Cross-border · e-Commerce · Consumer behavior · Consumer reviews · Regularized regression · Topic models

✉ Yang Li
yang.li@ruc.edu.cn

[1] Center for Applied Statistics, Renmin University of China, Beijing, China

[2] School of Statistics, Renmin University of China, Beijing, China

[3] School of Electrical Engineering and Computer Science, Peking University, Beijing, China

[4] Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong, China

[5] Statistical Consulting Center, Renmin University of China, Beijing, China

Published online: 23 July 2019

1 Introduction

With the development of Internet, China's e-Commerce industry has experienced explosive growth in the last decade. By 2016, China's total e-Commerce transactions reached 28.9 trillion yuan, an increase of 38.9% compared to 20.8 trillion in 2015 [32]. At the same time, China's cross-border e-Commerce also gradually grows. The cross-border e-Commerce import and export total 6.6 trillion yuan in 2016, according to statistics from China's Cross-Border E-commerce Research Center.

The rapid development of e-Commerce, especially cross-border e-Commerce, could not go further without the support of Chinese government. In recent years, Chinese government has paid great attentions to e-Commerce and played as an important role to its prosperity [32]. A land-mark event is China's Belt and Road Initiative (BRI) launched in 2016. This strategy seeks to improve infrastructure construction and deliver new technologies to countries beyond the immediate borders of China. The launch of BRI brings benefits to both China's economy and the world's economy [8, 28]. For China's economy, the Belt and Road Initiative can drive domestic development and strengthen China's economic collaboration with other countries. From the perspective of cross-border e-Commerce, the BRI policy opens new window and provides great opportunities for Chinese cross-border e-Commerce companies to go global. For the world's economy, the implementation of BRI brings infinite possibilities for countries along the Belt and Road. Until 2017, Chinese companies have invested more than 50 billion and built 56 economic and trade cooperation zones in 20 countries along the Belt and Road, which spurred the economic development of these countries to a great extend.

Challenges always go hand in hand with opportunities. How to seize the development opportunities and cope with the challenges is of great concern to cross-border e-Commerce companies. Although cross-border e-Commerce shares similarities with domestic launch and marketing strategies, it has substantial differences. There exists a large literature focusing on studying issues related to cross-border e-Commerce, such as business strategies [1, 7, 20], logistics service [15], cultural adaptation [2, 27], and so on. In this paper, we are interested in analyzing behaviors of overseas consumers. Consumers' online experiences are critical to website competitiveness [19]. Therefore, investigating consumers' online experiences provides an effective way to cognize consumers and thus create opportunities for cross-border e-Commerce companies.

In order to analyze consumer behaviors, online consumer reviews, cheap and abundant by its nature, provides an easy access. Numerous studies have verified the influences of online consumer reviews on the market success [6, 17, 21]. On one hand, potential consumers could rely on the re-views to shape product attitudes and make purchase decisions [24]. On the other hand, companies regard reviews as a new communication mechanism, and tend to promote new products through proactive management of the reviews [5, 12, 25]. Therefore, both researches and practitioners have great interest in investigating consumer reviews

and measuring their influence on products market performance [11]. In this paper, we focus on consumer reviews of a certain product (i.e., the cellphone), and try to find differences between behaviors of domestic and overseas consumers.

A consumer review usually consists of a numerical rating indicating the reviewers' overall attitude, and a text content describing the reviewers' evaluation in detail. Over the past few decades, studies of consumer reviews mainly focus on the rating information but barely explored the text content [6, 11, 17]. Recently, text-based analysis of consumer reviews has attracted considerable attention [4, 10, 33]. Since consumers' evaluations about products are often multi-dimensional, which cannot be captured by a single numerical rating, text-based analysis of reviews provides richer information for researchers and practitioners to understand consumers.

In this paper, we apply topic modeling techniques to analyze consumer reviews. Topic models are a suite of models that aim to discover and annotate large archives of documents with thematic information [3, 4, 13]. The basic topic model is the latent Dirichlet allocation (LDA) [3]. The key idea of this model is that each document offers a probability mixture over a common set of latent topics and each latent topic is a probability distribution over a dictionary of words. Though topic models, we are able to discover the latent topics underlying consumer reviews and thus make summarization for the whole review corpus. In recent years, topic models have been widely used to explore textual contents in management and marketing fields [14, 23, 26].

In this work, we apply the biterm topic model (BTM) [30], particularly designed for short texts, to extract topics from consumer reviews. We then evaluate the effects of topics on product rating scores. Results show that reviews of domestic consumers and overseas consumers contain different latent topics, which indicates different concerns of product. Specifically, the overseas consumers care more about product quality and tend to make description of technique details. On the contrary, domestic buyers pay more attention on consumer services and intend to comment in generalities. In addition, after accounting for other factors, we find certain topics are significantly associated with the product rating scores, with varied effects across reviews of domestic consumers and overseas consumers. All these findings could help e-Commerce companies design better launch strategies in cross-border e-Commerce market.

The rest of this paper is organized as follows. Section 2 introduces the research methodology. Section 3 describes the review dataset of cellphones and gives descriptive analysis results. Section 4 presents the empirical results of topic models and linear models. Section 5 concludes with a brief discussion.

2 Research methodology

2.1 Analyzing framework

We focus on consumer reviews of a specific product, the cellphones, to explore differences between behaviors of domestic and overseas consumers. To achieve this research goal, we take the following main steps. First, we collect cellphone

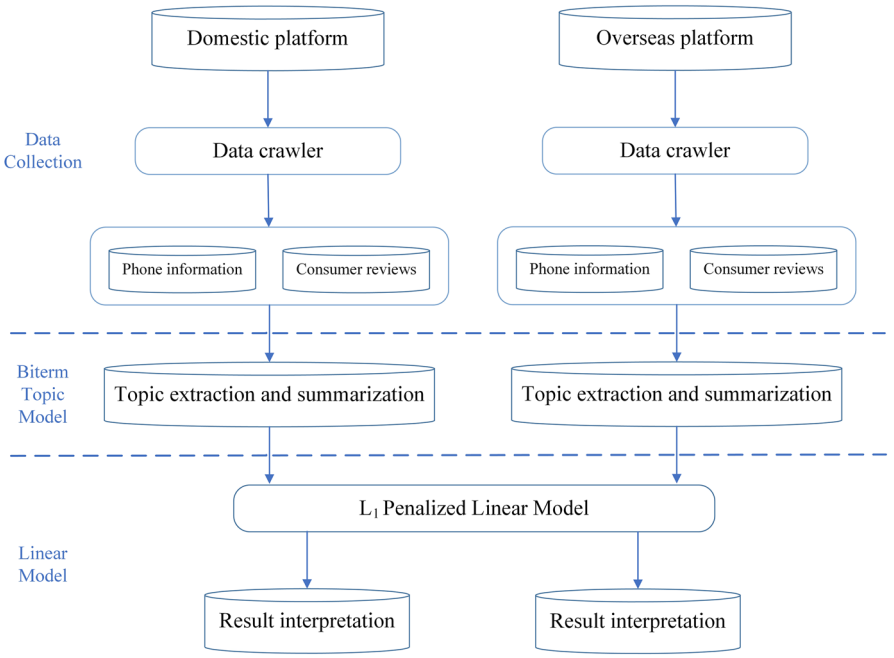

Fig. 1 The analyzing framework

information and consumer reviews from two sales platforms owned by one famous B2C online retailer in China. The two sales platforms are for domestic consumers and overseas consumers, respectively. Second, the biterm topic model is applied for the collection of consumer reviews. To better summarize the extracted topics, we classified all topics into six groups. Finally, we conduct linear models to explore the relationship between each topic group and the cellphone rating scores. The analyzing framework is present in Fig. 1.

Below, we will elaborate the methods we used in this work. For better illustration, we first summarize the key terms and mathematical symbols in Table 1. The detailed explanations are present in the following sections.

2.2 Topics extraction by using biterm topic model

Topic models are a popular paradigm of analyzing text documents, which can reveal the thematic structure in the document collection. Assume there are a total of D documents, which constructs a document corpus. All unique words appearing in this document corpus constitute a dictionary with size V. The basic assumption of topic models is that, there are K latent topics underlying the document corpus. Each document d is a mixture of these K topics according to a probability vector $\theta_d = \left(\theta_{d1}, \ldots, \theta_{dK}\right)^T$, which θ_{dK} represents the probability of topic k appearing in document d. Each topic k is characterized by its probability distribution

Table 1 Summary of key terms and mathematical symbols used in methodology

Key terms	
Corpus	The collection of all document
Dictionary	The collection of all unique words
Biterm	An unordered word-pair
Lasso	A variable selection method using L_1 penalized function
Mathematical symbols in topic models	
K	The number of topics
D	The number of documents
V	The number of unique words appearing in all documents
θ_{dk}	The probability of topic k appearing in document d
θ_d	The vector of topic probabilities in document d, i.e., $\left(\theta_{d1}, \ldots, \theta_{dK}\right)^T$
θ	The matrix of topic probabilities for all documents, i.e., $\theta = \left\{\theta_1, \ldots, \theta_N\right\}$
ϕ_{dk}	The probability of the vth word appearing in topic k
ϕ_d	The vector of word probabilities in topic k, i.e., $\left(\phi_{k1}, \ldots, \phi_{kV}\right)^T$
ϕ	The matrix of word probabilities for all topic, i.e., $\phi = \left\{\phi_1, \ldots, \phi_K\right\}$
$\left(\omega_i, \omega_j\right)$	A biterm constructed with words ω_i and ω_j
z_{ij}	The topic represented by $\left(\omega_i, \omega_j\right)$
$Dir(.)$	The Dirichlet distribution
$Multi(.)$	The multinomial distribution
Mathematical symbols in linear models	
N	The number of cellphones
p	The number of covariates
Y_i	The response variable for the ith product
X_{ij}	The value of the jth covariate related to the i cellphone
X_i	The vector of all covariates related to the $i\psi$cellphone, i.e., $\left(X_{i1}, \ldots, X_{ip}\right)^T$
β_0	The intercept of the linear model
β_j	The regression coefficient related to the jth covariate
β	The vector of all coefficients, i.e., $\left(\beta_1, \ldots, \beta_p\right)^T$

$\phi_k = \left(\phi_{k1}, \ldots, \phi_{kV}\right)^T$, where ϕ_{kV} represents the probability of the vth word appearing in topic k.

Practically, most of consumer reviews published on online sales platforms are short texts. When faced with a corpus of short texts, most topic models suffer from severe data sparsity problem [30, 31]. Specifically, the small word counts and short contents would restrict the ability of topic models to learn distinguished topics from the underlying text corpus. To address this problem, the biterm topic model (BTM) focus on "biterms", i.e. an unordered word-pair, and learn word co-occurrence on the whole corpus, not only on the document level [30]. The effectiveness of using BTM to handle short texts has been verified in many studies, such as [22, 34], and [35].

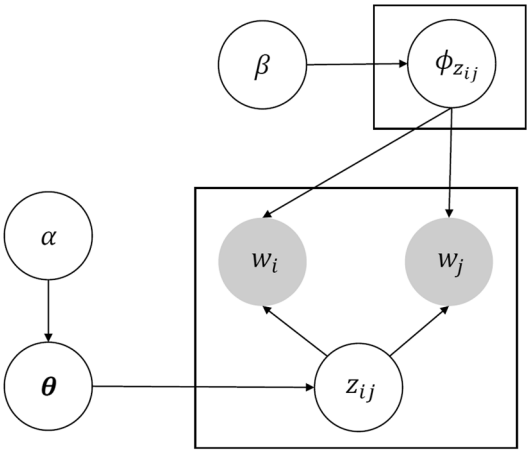

Fig. 2 The graphical representation of BTM

In BTM, a biterm denotes an unordered word-pair co-occurring in a short context (i.e. an instance of word co-occurrence pattern). Here, we regard any two distinct words in a consumer review as a biterm. For example, in the review "a great smart phone", if we ignoring the stop word "a", there are three biterms, i.e. "great smart", "great phone", and "smart phone". The biterms extracted from all short texts are collected as the training data of BTM. Then, BTM learns topics over the short texts based on the collection of biterms to solve the sparsity problem in a single text document. Specifically, the whole collection of text documents is assumed to have a probability vector $\boldsymbol{\theta}$ over K topics. Each topic k is assumed to have a probability vector $\boldsymbol{\phi}_k$ over a total of V words. Each biterm (ω_i, ω_j) only express one topic $z_{ij} \in \{1, \ldots, K\}$. Given z_{ij}, the two words ω_i and ω_j are assumed to be independently drawn from topic-related probability. The generative process of BTM is described as follows.

1. For each topic $k = 1, 2, \ldots, K$.
   (a) Draw a topic-specific word distribution $\boldsymbol{\phi}_k \sim Dir(\beta)$.
2. Draw a topic distribution $\boldsymbol{\theta} \sim Dir(\alpha)$ for the whole collection.
3. For each biterm $\mathrm{b} = \left(\omega_i, \omega_j\right)$.

   (a) Draw a topic indicator $z_{ij} \sim Multi(\boldsymbol{\theta})$.
   (b) Draw two words: $\omega_i, \omega_j \sim Multi\left(\boldsymbol{\phi}_{z_{ij}}\right)$.

Here, *Dir* denotes the Dirichlet distribution, *Multi* denotes the multinomial distribution, α and β are parameters in Dirichlet distributions. Figure 2 shows the graphical representation of BTM. One can see that, BTM directly models the word co-occurrence pattern (i.e. the biterm), rather than a single word. By doing so, the model is able to leverage the rich global word co-occurrence patterns to enhance the learning of topics. After model fitting of BTM, we could obtain the topic probabilities for each text document (i.e. $\boldsymbol{\theta}$) and word probabilities for each topic (i.e. $\boldsymbol{\phi}_{\mathrm{k}}$).

2.3 Topic influence exploration by using regularized linear models

We apply the linear model to investigate the influence of extracted topics on product rating scores. Assume there are a total of N cellphones. Let $Y_i (i = 1, \ldots, N)$ be the response variable for the ith product, and $X_i = \left(X_{i1}, \ldots, X_{ip}\right)$ be the corresponding vector of p covariates. Then, the model is specified as

$$Y_i = \beta_0 + X_i^T \beta + \varepsilon_i \tag{1}$$

where β_0 is the intercept, $\beta = \left(\beta_0, \ldots, \beta_p\right)^T$ is vector of coefficients related to X_i, and ε_i is the random noise with mean 0. Given the number of covariates could be large, we apply Lasso to select important covariates [29]. As a result, a L^1 penalized regression is established and the estimates of (β_0, β) are obtained as follows:

$$\left(\hat{\beta}_0, \hat{\beta}\right) = \arg\min_{\beta_0, \beta} \left\{ - \sum_{i=1}^{N} \left(Y_i - \beta_0 - X_i^T \beta\right)^2 + \lambda \sum_{j=0}^{p} \beta_j \right\}, \tag{2}$$

where λ_j is a tuning parameter. We follow [9] and [16] to use tenfold cross-validation to select λ_j that minimizes the prediction error. Except for Lasso, other feature selection methods, such as SCAD [18], Elastic Net [37] and Adaptive Lasso [36] are also considered to test the stability of selected features.

3 Data

3.1 Data collection

Cellphone information and consumer reviews from two sales platforms owned by Jingdong, one of the largest B2C online retailers in China, are collected for demonstration.[1] Specifically, from JD (www.jd.com), which sells products for Chinese consumers, we collected information of 897 cellphones and 384,359 corresponding consumer reviews. From Joybuy (www.joybuy.com), which sells products for overseas consumers, we collected information of 655 cellphones and 112,155 post-ed reviews. Noting some phone related features are inaccessible, we utilize information from two other websites (i.e., www.devicespecications.com and www.phonearena.com) to make up for the incomplete information. All information is crawled from the corresponding website by using the URLLBI and SELENIUM libraries in Python. After data collection, we removed reviews with < 5 words and then removed cellphones with < 10 reviews. As a result, we have 339 and 263 cellphones left for domestic platform and overseas platform, respectively.

All phone related features are listed in Table 2. On both platforms, each review is associated with a rating score, which varies from 1 to 5. Generally speaking, a good review refers the one with a rating score no lower than 4 and a bad review refers

[1] A data sample and related codes can be found in https://github.com/ffair/e-Commerce-analysis.

Table 2 Description of phone related features

	Variable	Description	Summary
Continuous variables	Rating score	Cellphone rating score	1–5
	Price	Cellphone price (RMB)	1250–8000
	Screen	Size of screen (in.)	2.4–6.6
	Weight	Cellphone weight (g)	60–385
	Thickness	Cellphone thickness (cm)	5.5–20
	Cam-Front	Pixels of front camera	190–2100
	Cam-Back	The pixels of back camera	200–2300
	ROM	The ROM storage (G)	2–128
Discrete variables	Network	OS version	3G, 4G
	Chip	The chip of cellphones	CPU, GPU
	GPS	Whether GPS is available	Support, not support
	Brand	Cellphone brand	Huawei, Xiaomi, et al.

the one with a rating score no higher than 2. Thus, the rating score of cell phone is defined as the ratio of good review among all review. In this work, we regard cellphone rating scores as the dependent variable. Except for rating scores, other cellphone related features would be taken into account as control variables.

With regard to consumer reviews, we first conduct text preprocessing before topic modeling. Following common practice in text mining, we first preprocessed the consumer reviews by using the NLTK library in Python to remove numbers and punctuations. For Chinese reviews in the domestic platform, we performed an additional processing step by segmenting Chinese sentences into word sequences using an open source package *jieba*. Afterward, we remove stop words, which are commonly used but have little semantic meaning in most occasions, such as "is" and "the". As a result, 81,062 unique Chinese words and 15,860 unique English words are left in each review corpus.

3.2 Descriptive analysis

We provide some descriptive analysis of the review corpus. The distribution of review length and post count by each consumer in Chinese corpus is displayed in Fig. 3. In Fig. 3a, we find that the distribution of review length is highly right-skewed, and most reviews have <50 words. The distribution of reviews posted by each consumer is shown in Fig. 3b. One can see that, this distribution follows the power law distribution with a heavy tail. It suggests that while most consumers make few comments on cellphones, there are a few consumers have large amount of purchases and comments. Similarly, the distributions of reviews from overseas consumers have similar patterns, i.e., the right-skewed distribution of review length shown in Fig. 4a and power-law distribution of review post count shown in Fig. 4b.

Lastly, we investigate the distributions of rating scores, which are regarded as dependent variables in the subsequent analysis. Figure 5 shows that, the distributions of ratings scores posted on both platforms are highly left skewed. These

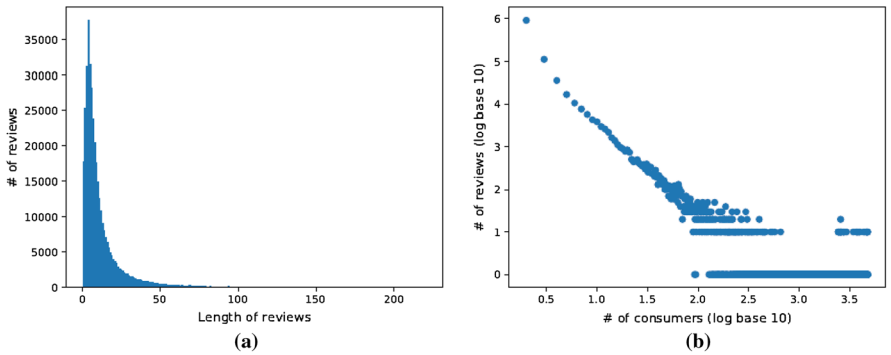

Fig. 3 The distribution of review length (**a**) and post count by each consumer (**b**) in domestic platform

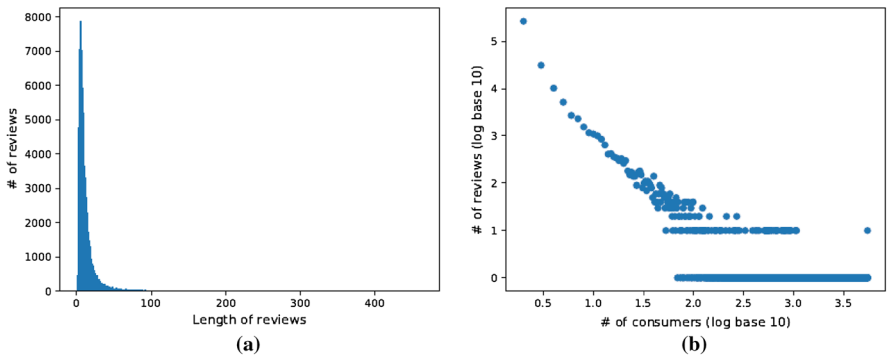

Fig. 4 The distribution of review length (**a**) and post count by each consumer (**b**) in overseas platform

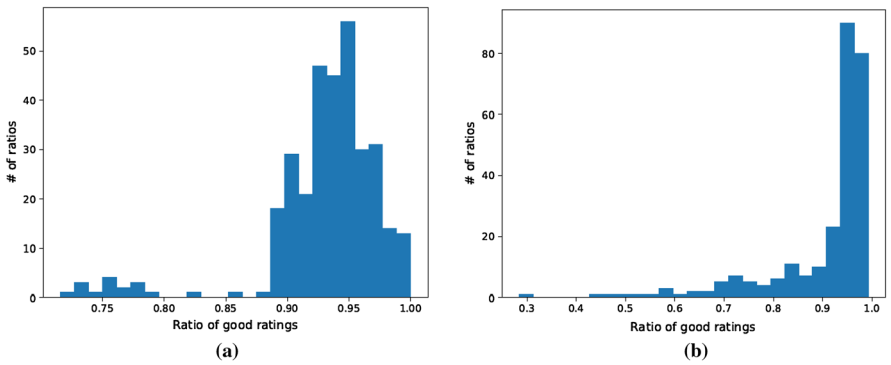

Fig. 5 The distribution of cellphone rating scores for reviews on **a** domestic platform, and **b** overseas platform

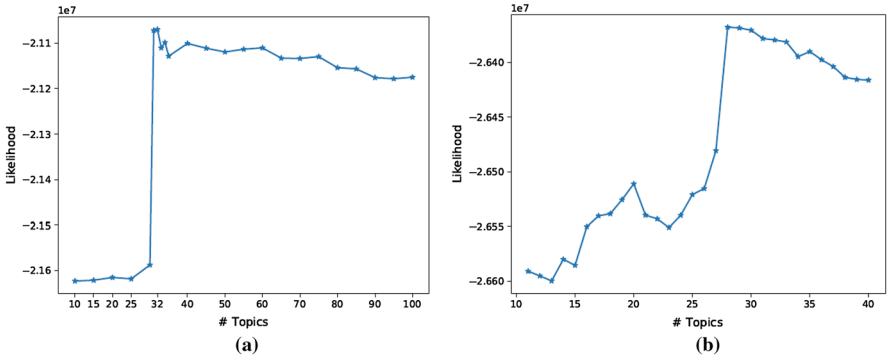

Fig. 6 The log-likelihood values under different number of topics for review collection in the domestic platform (**a**) and overseas platform (**b**), respectively

findings suggest that most cellphones have achieved relatively high rating scores. Specifically, the rating scores on the domestic platform ranges from 0.65 to 1.00, with 94.72% of the scores fall between 0.85 and 1.00. In the overseas platform, the rating scores fall into a wider interval, which ranges from 0.3 to 1.00.

4 Empirical results

4.1 Results of topics

We apply the biterm topic model to extract topics from consumer reviews collected in domestic platform and overseas platform, respectively. For reviews on each platform, it is assumed that there are K common topics for all reviews. Following common practice in biterm topic model, we set the hyper-parameters as $\alpha = 50/K$ and $\beta = 0.01$ [30]. To decide the optimal value of K for reviews on each platform, we let K vary from 10 to 100, and calculated the corresponding log-likelihood value for the model under each given K. Figure 6 shows the patterns of log-likelihood for each review dataset. Results show that the log-likelihood is the largest when $K=32$ for consumer reviews on the domestic platform, and when $K=28$ for consumer reviews on the overseas platform. Thus, we select 32 and 28 as the number of topics extracted from the two type of reviews, respectively.

After model fitting, we obtain the word probabilities $_{\mathrm{k}}$ for each topic k ($k = 1, \ldots, K$). Words with high probabilities under each topic are used to characterize the topic. Meanwhile, we get the topic probabilities $\boldsymbol{\theta}_d$ for each review d, where the kth element of $\boldsymbol{\theta}_d$ is the probability of topic k discussed in review d. For a better understanding, we label each topic based on the words with relatively high probabilities under that topic. All these topics are then categorized into separate groups based on thematic associations between them. Figure 7 presents the above process of applying biterm topic models to extract topics from consumer reviews on two platforms.

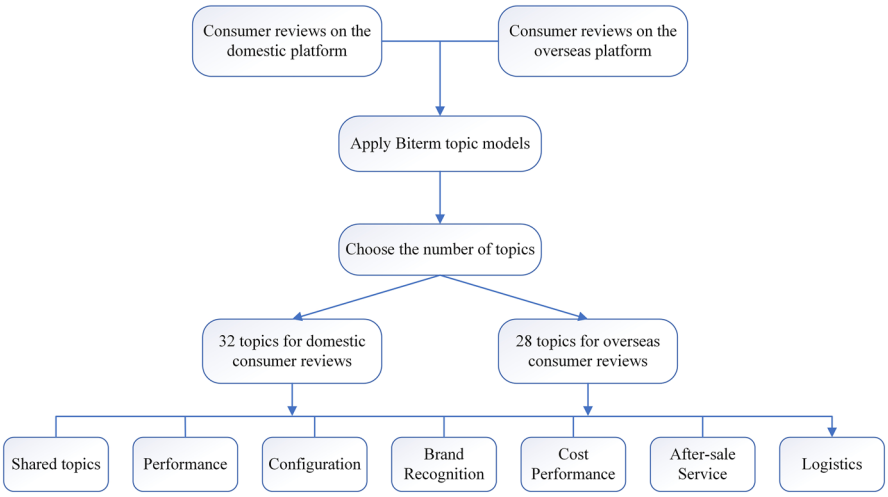

Fig. 7 The process of applying biterm topic models to extract topics from consumer reviews on two platforms

Table 3 lists the extracted topics from reviews on the domestic platform. As one can see, most topics underlying reviews on the domestic platform are meaningful. However, there are also topics, which do not concentrate on a specific meaning, but mainly comprise words shared by all the other topics. We refer this kind of topics as *shared* topics. Except for the shared topics, the other topics are classified into six categories. They are, respectively, *Performance*, *Configuration*, *Brand Recognition*, *Cost Performance*, *After-sale Service* and *Logistics*.

Table 4 presents the extracted topics from reviews on the overseas platform. Similar with results shown in Table 3, topics extracted from overseas reviews are also classified into the six categories. However, even in the same topic category, topics extracted from two platforms have differences. Take the category *Performance* as an example. As shown in Tables 3 and 4, both domestic consumers and overseas consumers care about the OS performance, battery duration, signal strength and system fluency when playing games. In addition, domestic consumers also pay more attention to screen display and call quality. As for overseas consumers, they tend to make more comments on technique details, such as touch sensitiveness and battery fervescence. The detailed information of extracted topics is present in "Appendix 1".

4.2 Topic comparison

We then investigate the probabilities of the extracted topics from consumer reviews at the domestic platform and the overseas platform, which can help us better understand the content differences between the two type of reviews. To focus on the meaningful content (the content containing non-shared topics) when examining content differences between consumer reviews on two platforms.

Table 3 Topic categories, topic number, topic names, and example words from reviews on the domestic platform. All Chinese words are translated into English

Topic category	Number	Topic name	Example words with high probabilities
Performance	1	Overall	Running, smoothly, outlook, screen, taking pictures
	2	OS Usage	OS, function, update, software, operating
	3	Gaming	Play, games, Glory, not stuck, smooth
	4	Screen display	Screen, brightness, colors, display, show o_
	5	Battery duration	Hour, battery, fully charged, standby, durable
	6	Signal	Signal, bad, network, connections, full signal
	7	Call quality	Sound, hear, receiver, noise, voice
	8	OS optimization	Software, OS, RAM, install, power on
Configuration	9	Sim card	Card, Unicom, Telecom, card slot, insert
	10	Screen	Screen, break, fall down, quality, glass
	11	Basic function	Configure, alarm, call, display, bluetooth
	12	Native APPs	Weichat, download, photos, data, games
	13	Shell and film	Film, shell, buy, tempering, earphone
Brand recognition	14	Overall	Hammer, Hope, Meizu, design, Xiaomi
	15	Homemade	Huawei, Xiaomi, Homemade, Honor, brand
	16	The outdated	Nokia, made, Blackberry, old fashion, disappoint
	17	Joybuy	Joybuy, goods, purchase, reliable, satisfy
Cost performance	18	Cost performance	High, cost performance, price, buy, cheap
	19	Preferential activity	Price, order, purchase, promotion, interest Free
	20	Price concessions	Buy, price reduction, price insured, reduce, coupon
After-sale service	21	Repairing	After-sale, repair, detection, broke, request
	22	Returning	Quality, not good, suck, regret, repair
	23	Comments	Negative comments, one star, like, too bad, positive
	24	Third-party seller	Seller, satisfy, 3rd party, package, professional
Logistics	25	Logistics dispatch	Goods delivery, recipient, package, order, soon
	26	Sealing off	Scratch, open, exchange, take over, 2nd hand
	27	Jingdong logistics	Joybuy, delivery, speed, service attitude, fast
	28	Packaging	Package, box, open, rough, logistics
Shared	29	Shared	Multiple, found, landslide, mine, warmth
	30	Shared	Comment, user, content, default, word
	31	Shared	Bit, little, issue, thick, heavy
	32	Shared	Http, jpg, info, watch, use

To make the subsequent analyses more manageable and interpretable, we only focus on the meaningful content containing non-shared topics when examining content differences between consumer reviews on two platforms. Thus, we do not take shared topics into consideration. To remove the influence caused by shared topics, we conduct a normalization procedure. Specifically, let θ_{dk}^{type} denote the probability of topic k discussed in review d posted on the given type (type = "d" for domestic

Table 4 Topic categories, topic number, topic names, and example words from reviews on the overseas platform

Topic category	Number	Topic name	Example words with high probabilities
Performance	1	Overall	Good, phone, smooth, fast, outlook
	2	Picture taking	Screen, camera, picture, clear, photo
	3	Outlooking	Beautiful, look, color, feel, appearance
	4	Signal	Signal, bad, mobile, network, wifi
	5	Gaming	Play, game, card, Glory, smooth
	6	Battery duration	Battery, charge, power, durable, charging
	7	Photographing	Picture taking, photograph, pixel, camera, clear
	8	OS failure	Crash, shut down, OS, stuck, open up
	9	Touch sensitiveness	Screen, key, button, fingerprint, insensitive
	10	Fervescence	Heating, hot, generate fever, ironing, explode
Configuration	11	Hardwares	System, software, memory, instal, upgrade
	12	Material and feel	Screen, glass, cover, metal, touch
	13	Acoustic	Sound, voice, listen, tone, music
	14	Network	Mobile, telecom, unicom, communication, netcom
	15	Unlock	Version, unlock, global, international, problem
Brand recognition	16	Jingdong	Jingdong, website, shop, trust, mall
	17	Domestic	Huawei, national, domestic, Xiaomi, Meizu
	18	Comparison	Meizu, Xiaomi, better, HTC, Blackberry
Cost performance	19	Cost performance	Worth, buy, quality, great, value
	20	Cost performance	High, value, price, cost, money
	21	Evaluation and comparison	Evaluate, product, copy, compare, difference
After-sale service	22	Customer service	Customer, service, buy, complaint, solve
	23	Returns policy	Return, customer, exchange, purchase, refund
Logistics	24	Packaging and delivery	Package, delivery, box, order, fast
	25	Delivery experience	Delivery, receive, logistics, satisfied, distribution
	26	Delivery process	Express, receipt, deliver, send, locate
Shared	27	Shared	One, use, phone, year, plus
	28	Shared	Word, sleep, one, dimensional, code

and type = "o" for overseas) of platform. Then, we normalize θ_{dk}^{type} to p_{dk}^{type}, so that the sum of $p_{dk}^{(type)}$ for all non-shared topics equals 1. That is,

$$p_{dk}^{type} = \theta_{dk}^{type} / \sum_{k \in \Lambda^{type}} \theta_{dk}^{type}, \tag{3}$$

where Λ^{type} is the set of all non-shared topics. After normalization, we calculate the total probability of topics in each category, i.e.

Table 5 Results of two-tail *t*-tests of category probability for domestic reviews and overseas reviews

Topic category	Average probability in domestic reviews	Average probability in overseas reviews	*P* value
Performance	0.164	0.228	< 0.001
Configuration	0.153	0.241	< 0.001
Brand recognition	0.095	0.155	< 0.001
Cost performance	0.132	0.137	0.107
After-sale service	0.239	0.097	< 0.001
Logistics	0.217	0.142	< 0.001

$$p_{dk}^{type} = \sum_{k \in \Lambda^{type}} p_{dk}^{type}, \tag{4}$$

where Λ_g^{type} $(g = 1, \ldots, 6)$ denotes the set of topics in *Performance*, *Configuration*, *Brand Recognition*, *Cost Performance*, *After-sale Service* and *Logistics*, respectively.

To examine the content differences between the two types of reviews, we perform a series of two-tail *t*-tests on the probabilities of each category. Results of *t*-tests (shown in Table 5) indicate that, domestic reviews discuss more on topics in *After-sale Service* and *Logistics*; while overseas consumers discuss more on topics in *Performance* and *Configuration*. In addition, overseas consumers pay more attention to product brands than domestic consumers. As for the category *Cost Performance*, the two type of consumers do no show significant differences. These findings imply that, while it is hard for overseas consumers to get in touch with cellphone companies, they focus more on product quality and brands, but less on customer services.

4.3 Regression results

To explore the impact of extracted topics on consumers' evaluations, we conduct the linear regression models for cellphone datasets collected on domestic platform and overseas platform, respectively. The models are specified by incorporating the logarithm of cellphone rating scores as the dependent variable and topic probabilities in each category as independent variables, along with all the phone related variables and service-related variables (see Table 1 for details) as control variables. Given the number of variables is large, we apply Lasso [29] for variable selection. After using Lasso, a linear regression model only with selected variables is built to obtain unbiased coefficients.

Regression results for two cellphone datasets collected on domestic platform and overseas plat-form are shown in Table 6. It is obvious that, all variables left in the two regression models are significant. Specifically, for the domestic dataset, the price, brand Huawei, OPPO, and the screen all have positive impacts on cellphone rating scores. Regarding the topic probabilities, four categories of topics are selected

Table 6 Results of linear regression model with variables selected by Lasso

Variable	Coefficient	*P* value	Variable	Coefficient	*P* value
Intercept	−0.177	0.012	Intercept	−0.627	0.004
Price	0.251	< 0.001	Price	0.353	< 0.001
Huawei	0.920	< 0.001	Huawei	0.845	0.001
OPPO	0.337	0.051	Xiaomi	0.466	0.047
Screen	0.423	0.038	Back camera	0.574	0.038
Performance	−0.188	0.009	Front camera	0.586	0.022
Configuration	−0.161	0.005	Performance	−0.224	0.003
After-sale service	−0.249	0.001	Configuration	−0.259	0.033
Logistics	0.169	0.030	Logistics	−0.226	0.004
Adjusted R^2	51.4%		Adjusted R^2	56.2%	

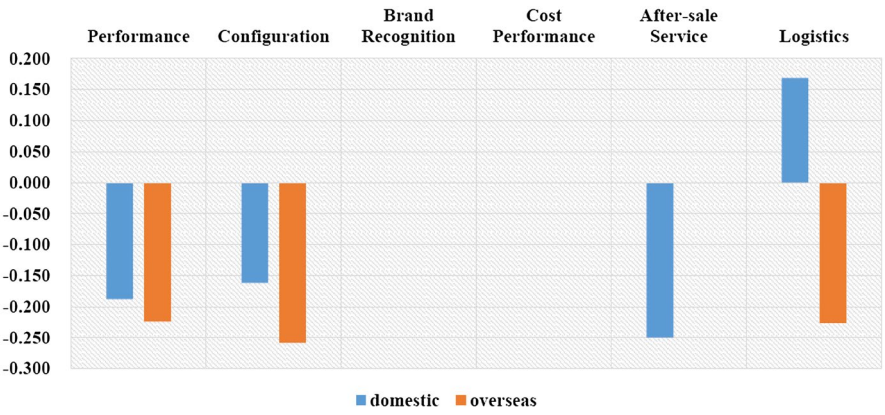

Fig. 8 The bar plot of coefficients associated with topic categories

by the model, i.e. *Performance*, *Configuration*, *After-sale Service* and *Logistics*. For the overseas platform, the overseas dataset, the price, brand Huawei, Xiaomi, as well as camera pixels all play positive roles on the cellphone rating scores. In terms of topic probabilities, only three categories of topics are left. They are, respectively, *Performance*, *Configuration*, and *Logistics*.

Except for Lasso, we also try other feature selection methods, such as SCAD [18], Elastic Net [37] and Adaptive Lasso [36]. The detailed selection results regarding the six topic categories under different feature selection methods are present in "Appendix 2", which are similar with those under Lasso. Thus, we focus on the selection results obtained by Lasso in the subsequent analysis. To better illustrate the impacts of topic probabilities, we draw a barplot to compare the two groups of coefficients associated with topic categories. The barplot in Fig. 8 shows that, topic categories *Brand Recognition* and *Cost Performance* have no effects to cellphone rating scores in both platforms. Note that both regression models have included price and brands as control variables, the two topic categories discussed in reviews have no

more information that can explain rating scores. Regarding the categories *Performance* and *Configuration*, they both have negative effects to rating scores on the domestic platform and overseas platform. However, the overseas platform has higher absolute values of coefficients. These findings suggest that, the cellphone performance and configuration are pain points to both domestic consumer and overseas consumers, which are even worse to the latter.

The topic category *After-sale Service* has a significant negative effect on rating scores on the domestic platform, indicating domestic consumers are not satisfied with the after-sale service. However, this topic category shows no effect to those on the overseas platform. A possible explanation might be that, overseas consumers have relatively low expectations on the after-sale service of cross-border e-Commerce. Lastly, the topic category *Logistics* has opposite effects on rating scores on the two platforms. Specifically, domestic consumers are satisfied with the logistics speed, but overseas consumers are disappointed.

In summary, for domestic consumers, four aspects, i.e., the cellphone performance, configuration, after-sale service and logistics, could significantly influence customer evaluations. As for overseas consumers, the cellphone performance, configuration and logistics, are significantly related to consumers evaluations. These findings can provide managerial implications for policy makers of cross-border e-Commerce to better adapt to the overseas market. Firstly, both domestic consumers and overseas consumers are not satisfied with cellphone performance and configurations. Thus, cellphone companies need to design well-behaved products in performance and configuration to improve the current impressions. Secondly, logistics is a positive factor on domestic platform, but a negative factor on overseas platform. Thus, securing fast logistics is the key to enlarge the overseas market. Finally, after-sale service plays a negative role on consumer shopping experiences on the domestic platform. Therefore, domestic e-Commerce platform needs to enhance its after-sale service quality to make consumers satisfied.

5 Conclusion and discussion

China's Belt and Road Initiative (BRI) policy has provided great opportunities for Chinese e-Commerce companies to sell goods overseas. While cross-border e-Commerce shares similarities of launch and marketing strategies with domestic e-Commerce, substantial differences still exist. The key difference lies in the fact that consumer behaviors on domestic and overseas shopping platforms are quite different. Therefore, how to make strategic adjustments to better adapt to the overseas market is of great concern to cross-border e-Commerce companies.

In this work, we conduct comparison between behaviors of domestic consumers and overseas consumers, through their posted reviews. We focus on a specific

product, i.e. cellphones, and apply topic modeling techniques to extract richer information from review contents. By using the biterm topic model, which particularly designed for short texts, we obtain six categories of topics underlying reviews posted by domestic and overseas consumers, respectively. Among these topic categories, domestic consumers pay more attention on consumer services (*After-sale Service* and *Logistics*); while overseas consumers focus more on cellphone quality and brands (*Performance*, *Configuration* and *Brand Recognition*). Lastly, results of linear regression models demonstrate the varying effects of topic categories to consumer evaluations (i.e., the cellphone ratings) across the domestic platform and overseas platforms. These findings provide strategic suggestions for e-Commerce companies.

- On both domestic and overseas platform, cellphone performance and configurations play negative roles on consumer shopping experiences. Thus, cellphone companies need to design well-behaved products in performance and configuration.
- Logistics is the strength to domestic platform, which can be maintained in the future. However, it is a negative factor on overseas platform, and needs to be improved.
- After-sale service is a concern factor only to domestic consumers, but has no effect on overseas consumers. Thus, domestic e-Commerce platform needs to enhance its after-sale service quality to make consumers satisfied.

To conclude this article, we discuss here a few directions for further study. Firstly, the performance of BTM can be influenced by the setting of hyper-parameters, which is one limitation of topic modeling techniques. Thus, it is important to check the influence of hyper-parameter settings on the main findings in future works. Secondly, more feature selection methods can be used to test the impact of topics to consumer evaluations. Thirdly, more advanced models, such as hierarchical linear models, could be applied to model both domestic and overseas datasets in a unified frame-work. Lastly, the analysis framework via topic modeling of consumer reviews can be extended to more products.

Acknowledgements This work was supported by fund for building world-class universities (disciplines) of Renmin University of China, China Postdoctoral Science Foundation (No. 2017M18304), the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (No. 18XNLG02), the National Natural Science Foundation of China (No. 71771211). This work described in this paper was partial supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. 11507817).

Appendix 1: Detailed information of topics

See Tables 7 and 8.

Table 7 The extracted topics and example words from reviews on the domestic platform. All Chinese words are translated into English

Topics	Top fifteen words with high probabilities
1	Phone, not bad, running, smoothly, outlook, hold-feeling, feel, screen, picture taking, like, OS satisfied, goodlooking, cost performance, clear
2	OS, phone, function, update, nice, smoothly, feel, software, operate, Android, upgrade, optimize, MIUI, Apple, support
3	Play, kings, game, Glory, phone, chicken, game playing, stuck, smooth, pixel, fever, No stuck, operation, stuck, screen
4	Screen, clear, phone, feeling, bright, colors, display, colors, eyes, good, video, comfort, mode, show O, resolution
5	Electric, phone, hour, battery, battery-vol, fully charged, over night, standby, run out, power consuming, long lasting, surplus, power on, durable
6	Signal, phone, bad, network, not good, stuck, places, connections, buy, Internet surfing, phone call, stable, signal, Huawei, full signal
7	Sound, phone, hear, phone call, receiver, not good, video, volume, noise, voice, speak, Ringtone, hands-free, acoustic
8	Software, OS, RAM, phone, running, native, cleaning, backstage, install, advertisement, power on, install, smoothly, update, take up
9	Card, phone, dual cards, unicom, telecom, dual standby, Netcom, card slot, phone card, signal, RAM, insert, function, two pieces, network
10	Phone, screen, break, fall down, buy, drop, screen, quality, ground, glass, less than, carefull, film, after-sale, shell
11	Phone, configure, alarm, call, display, SMS, screen lock, message, call reminder, Ringtone, bluetooth, clock, password, calling, contacts
12	Phone, Wechat, software, download, stuck, display, pictures, customer service, function, photo, system, network, data, voice, games
13	Phone, film, present, shell, buy, tempering, earphone, stick, sticker, screen, nice, protect, phone cover, negative comment, original binding
14	Phone, hammer, product, Hope, real, Mr. Luo, Meizu, design, Xiaomi, Nuts, feelings, like, OS, experience, support
15	Huawei, phone, support, Xiaomi, homemade, hopefull, buy, Honor, Apple, brand, disappoint, homemade goods, Nubia, more and more, Samsung
16	Nokia, phone, feel, workmanship, product-qual, function, Blackberry, brand, key, old way, disappoint, Philips, durable, look for, old fashion
17	Joybuy, buy, goods, gad, purchase, phone, deserve, logistics, reliable, service, support, express, satisfy, speed, mall
18	Phone, high, cost, nice, price, buy, configuration, Xiaomi, price, screen, buy, deserve, outlooking, Huawei, cheap
19	Joybuy, present, customer service, buy, gift, earphone, activities, phone, yuan, price, place order, purchase, ask, sales promotion, interest free
20	Buy, price reduction, price, phone, insured, apply, yuan, reduce, bought, day, protect, less, price, comment, coupon
21	Phone, after-sale, repair, Joybuy, customer service, detection, broke, screen, day, quality, request, exchange, repair, guarantee, one month
22	Phone, rubbish, Joybuy, bad, negative comments, less than, screen, quality, not good, stuck, regret, Xiaomi, Huawei, not good, return, repair
23	Negative comments, bad, star, one star, not like, one, too bad, delivery, positive comments, two, three, comment, super, logistics, purchase

Table 7 (continued)

Topics	Top fifteen words with high probabilities
24	Goods, seller, satisfy, attitude, receive, buy, 3rd party, package, soon, answer, goods, enthusiasm, answer question, professional, pleasant surprise
25	Goods delivery, Joybuy, customer service, mobile phone, month, recipient, package, place order, exhibit, timing
26	Phone, screen, noticed, receive, scratch, open up, feel, rejected, exchange, sign of usage, goods exchange, first, take over, package, 2nd hand
27	Joybuy, delivery, goods, dispatch, products, speed, price, equal, performance, best, comment, Hope, service, play
28	Phone, package, express, Joybuy, open, goods, packing box, inside, rough, packing, box, receive, feeling, logistics
29	Sucks, Huawei, battery, phone, elevator, Apple, drag, good, signal, workmate, multiple, found, landslide, warmth
30	Comment, user, content, unfilled, praise, default, earned points, accumulate, word, seller, waste, many, on purpose
31	Bit, hot, little, hair, issue, Ipad, App, expensive, best, thick, slow, game, heavy, crank, orbit
32	Http, rabbit, jpg, info, img, dog, com, www, collapse, Youtube, watch, ZTE, save, rope, use

Table 8 The extracted topics and example words from reviews on the overseas platform

Topics	Top fifteen words with high probabilities
1	Good, use, hand, phone, screen, mobile, feel, smooth, fast, outlook, beautiful, machine, fingerprint, operation, battery
2	Screen, camera, picture, phone, effect, clear, high, photo, better, system, quality, color, pixel, function, smooth
3	Beautiful, look, color, hand, feel, super, big, black, appearance, gold, high, nice, light, red, thin
4	Phone, signal, good, problem, cell, screen, time, bit, bad, mobile, network, wifi, LTE, call
5	Play, game, card, memory, big, enough, time, screen, cell, mobile, glory, run, smooth, hot, battery
6	Battery, charge, day, fast, electricity, power, hour, durable, full, hot, heat, electric, time, charging, consumption
7	Picture taking, cam-lens, photograph, pixel, effect, phone, camera, pictures, front-cam, thousands of, clear, function, rear-cam, video, good
8	Phone, Reboot, screen, circumstance, crash, shut down, boot, OS, blank screen, stuck, month, keyboard, phone call, discover, open up
9	Screen, key, phone, return, button, feeling, operation, volume, bad, function, virtual, fingerprint, set up, sometimes, insensitive
10	Phone, heating, score, hot, running, generate fever, ironing, game play, Antutu, severe, Xiaomi, temperature, game, performance, explode
11	System, phone, software, screen, card, memory, application, update, fingerprint, Android, machine, keyboard, instal, camera, upgrade
12	Screen, film, good, back, shell, glass, cover, metal, edge, protective, touch, front, color, camera, frame
13	Sound, voice, listen, tone, music, power, earphone, headset, radio, speaker, bluetooth, interface, workmanship, song, headphone

Table 8 (continued)

Topics	Top fifteen words with high probabilities
14	Generation, mobile, card, Telecom, unicom, phone, communication, memory, standby, netcom, slot, network, signal
15	Chinese, phone, version, rom, Russian, unlock, global, loader, application, update, international, problem, Android
16	Jingdong, worth, believe, website, shop, trust, brand, genuine, satisfied ed, mall, real, recommend, prefer, assure, guarantee
17	Support, Huawei, national, brand, domestic, glory, quality, product, Xiaomi, trust, ZTE, performance, China, Meizu, Nubian
18	Good, Meizu, system, millet, red, Xiaomi, mobile, rice, machine, better, really, like, cell, HTC, Blackberry
19	Good, phone, worth, use, buy, cell, quality, great, fast, feeling, nice, value, support, money, enough
20	High, value, price, performance, cost, buy, money, worth, quality, configuration, effective, low, appearance, beautiful, screen
21	Evaluate, product, commodity, unevaluated, point, copy, compare, definitely, explain, buy, difference, problem, consumer, bad, go
22	Jingdong, customer, service, buy, phone, hear, goods, buy, experience, complaint, products, commodity, seller, disappointed, solve
23	Phone, return, activation, Jingdong, reason, customer, apply, buy, days, sale, exchange, purchase, refund, service
24	Day, package, delivery, box, order, time, great, week, month, send, long, track, fast, pack, mail
25	Jingdong, fast, delivery, service, receive, order, praise, logistics, satisfied, distribution, attitude, sale, send, customer, experience
26	Express, receipt, custom, service, phone, call, distribution, deliver, send, to, goods, logistics, service, find, locate, deliverer, sign, for
27	Buy, one, use, phone, year, Nubian, good, better, plus, time, mobile, system, first, mini, feel
28	Word, sleep, fifteen, one, dimensional, two, code, handsome, phone, scan, write, comment, ingenious, neat, admire

Appendix 2: Selection results regarding six topic categories

We consider different variable selection methods. The results regarding six topic categories are shown in the following table. It is clearly that, the topic categories selected by Lasso are the overlaps of all methods, implying the stability of these selected topic categories. Therefore, we focus on the results of Lasso in this work (see Table 9).

Table 9 Selection results regarding six topic categories under different selection methods

Method	Performance	Configuration	Brand recognition	Cost performance	After-sale service	Logistics
Domestic platform						
Lasso	√	√			√	√
SCAD	√	√	√		√	√
Elastic Net	√	√			√	√
Adap-Lasso	√	√		√	√	√
Overseas platform						
Lasso	√	√				√
SCAD	√	√			√	√
Elastic Net	√	√				√
Adap-Lasso	√	√		√		√

References

1. Asosheh, A., Shahidi-Nejad, H., & Khodkari, H. (2012). A model of a localized cross-border e-Commerce. *iBusiness, 4*(2), 136.
2. Bhagat, R. S., Kedia, B. L., Harveston, P. D., & Triandis, H. C. (2002). Cultural variations in the cross-border transfer of organizational knowledge: An integrative framework. *Academy of Management Review, 27*(2), 204–221.
3. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research, 3,* 993–1022.
4. Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science, 35*(6), 953–975.
5. Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science, 54*(3), 477–491.
6. Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*(3), 345–354.
7. Damanpour, F., & Damanpour, J. A. (2001). E-business e-commerce evolution: Perspective and strategy. *Managerial Finance, 27*(7), 16–33.
8. Fan, Z. (2018). Chinas belt and road initiative: A preliminary quantitative assessment. *Journal of Asian Economics, 55,* 84–92.
9. Genkin, A., Lewis, D. D., & Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics, 49,* 291–304.
10. Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science, 31*(3), 493–520.
11. Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science, 23*(4), 545–560.
12. Godes, D., Mayzlin, D., Chen, Y., et al. (2005). The Firm's management of social interactions. *Marketing Letters, 16*(3/4), 415–428.
13. Griffths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America, 101*(1), 5228–5235.
14. Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent Dirichlet allocation. *Tourism Management, 59,* 467–483.
15. Hsiao, Y., Chen, M., & Liao, W. (2017). Logistics service design for cross-border e-commerce using Kansei engineering with text-mining-based online content analysis. *Telematics and Informatics, 34*(4), 284–302.

16. Ifrim, G., Bakir, G., & Weikum, G. (2008). Fast logistic regression for text categorization with variable-length N-grams. In *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 354–362). New York.
17. Karniouchina, E. V. (2011). Impact of star and movie buzz on motion picture distribution and box office revenue. *International Journal of Research in Marketing, 28*(1), 62–74.
18. Kim, Y., Choi, H., & Oh, H.-S. (2008). Smoothly clipped absolute deviation on high dimensions. *Journal of the American Statistical Association, 103*(484), 1665–1673.
19. Kotha, S., Rajgopal, S., & Venkatachalam, M. (2004). The role of online buying experience as a competitive advantage: Evidence from third-party ratings for e-Commerce firms. *Journal of Business, 77*(2), 109–133.
20. Lefebvre, E., Lefebvre, L. A., Le Hen, G., & Mendgen, R. (2006). Cross-border e-collaboration for new product development in the automotive industry. In *Proceedings of the 39th annual Hawaii international conference* (Vol. 1, pp. 82–90).
21. Legoux, R., Larocque, D., Laporte, S., Belmati, S., & Boquet, T. (2016). The effect of critical reviews on exhibitors' decisions: Do reviews affect the survival of a movie on screen? *International Journal of Research in Marketing, 33*(2), 357–374.
22. Li, X., Zhang, A., Li, C., Guo, L., Wang, W., & Ouyang, J. (2018). Relational biterm topic model: Short-text topic modeling using word embeddings. *The Computer Journal, 62*(3), 359–372.
23. Lin, K. P., Shen, C. Y., Chang, T. L., & Chang, T. M. (2017). A consumer review-driven recommender service for web e-commerce. In *IEEE 10th conference on service oriented computing and applications (SOCA)* (Vol. 1, pp. 206–210).
24. Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing, 70*(3), 74–89.
25. Mayzlin, D. (2006). Promotional chat on the internet. *Marketing Science, 25*(2), 155–163.
26. Puranam, D., Narayan, V., & Kadiyali, V. (2017). The effect of calorie posting regulation on consumer opinion: A flexible latent Dirichlet allocation model with informative priors. *Marketing Science, 36*(5), 726–746.
27. Sinkovics, R. R., Mo, Y., & Hossinger, M. (2007). Cultural adaptation in cross border e-Commerce: A study of German companies. *Journal of Electronic Commerce Research, 8*(4), 109–133.
28. Tian, W. L. (2017). The "Belt and Road" initiative: A Chinese concept for global development. *Contemporary International Relations, 27*(4), 1–20.
29. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 7*(2), 267–288.
30. Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. In *Proceedings of the 22nd International conference on world wide web* (pp. 1445–1456).
31. Yang, Y., Wang, F., Zhang, J., Xu, J., & Philip, S. Y. (2018). A topic model for co-occurring normal documents and short texts. *World Wide Web Journal, 21*(2), 487–513.
32. Yue, H. (2017). *National report on e-Commerce development in China*. Inclusive and Sustainable Industrial Development Working Paper Series, Vol. 17, pp. 1–42.
33. Zhao, Y., Yang, S., Narayan, V., & Zhao, Y. (2013). Modeling consumer learning from online product reviews. *Marketing Science, 32*(1), 153–169.
34. Zhou, X., Ouyang, J., & Li, X. (2018). Two time-efficient Gibbs sampling inference algorithms for biterm topic model. *Applied Intelligence, 48*(3), 730–754.
35. Zhu, Q., Feng, Z., & Li, X. (2018). GraphBTM: Graph enhanced autoencoded variational inference for biterm topic model. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 4663–4672).
36. Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association, 101*(476), 1418–1429.
37. Zou, H., & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of Statistics, 37*(4), 1733–1751.