Comparative Study on Chinese Consumer Concerns of Thai and Foreign Cosmetics in the Context of Cross-border e-Commerce by LDA and Sentiment Analysis

Yuting Feng, Nopasit Chakpitak, Piang-or Loahavilai, Tirapot Chandarasupsang, and Lei Mu

Abstract—Cross-border e-commerce has become one of the important trade models nowadays, especially for the countries under China's "The Belt and Road Initiative" policy, and Thailand has the strongest cross-border e-commerce connection with China among these countries. Cosmetics is the most popular category on Tmall Global which is one of the largest cross-border e-commerce retail platforms in China. However, Thai cosmetics sales are far behind top three sales countries which are Japan, South Korea and France. Electronic word of mouth has an important impact on Chinese consumer online purchase decision, and massive online reviews reflect consumers' concerns. In order to understand the reasons that Chinese consumer prefer cosmetics more from Japan, South Korea and France rather than Thailand, based on the data collected from China cross-border e-commerce industry and massive literature review, in this paper, two research problems are identified as consumer concern factors and their sentiment preference. Then 25 sample products and reviews are selected from these four countries by keywords searching, the relationship between sales volume and online reviews analysis indicate Chinese consumer has strong concerns and express their positive and negative sentiment of products in each country, especially Thai products. What is more, 9900 reviews are crawled in each country as raw input data. Finally, a text-mining based conceptual framework is introduced to apply in cross-border e-commerce cosmetics context which including latent Dirichlet allocation and sentiment analysis. The extracted factor and their proportion in each country, the factor difference comparison among four countries and sentiment preference of each factor are output after implement in the future. This research can benefit for products supplier, platform operator, logistics provider, researchers and so on, especially for participants along "The Belt and Road Initiative" countries. It can indicate the Chinese customer concerns gap between countries along "The Belt and Road Initiative" and other popular cosmetics countries.

Index Terms—Consumer concerns, cross-border e-commerce, electronic word of mouth, sentiment analysis, text-mining conceptual framework, latent Dirichlet allocation.

Yuting Feng, Nopasit Chakpitak, Piang-or Loahavilai, and Tirapot Chandarasupsang are with the International College of Digital Innovation, Chiang Mai University, Thailand (e-mail: elenafeng1993@gmail.com, nopasit@cmuic.net, piangor@gmail.com, tirapot@gmail.com).

Lei Mu is with the College of Foreign Languages and Cultures, Chengdu University, China (e-mail: mulei@cdu.edu.cn).

I. INTRODUCTION

A. Cross-border e-Commerce and "The Belt and Road Initiative"

Cross-border e-commerce (CBEC) refers to a new type of international business activity, which is a transaction entity that belongs to different countries. It can deal through e-commerce platform and use cross border logistics to distribute commodities [1].

CBEC includes cross-border import and export ecommerce. China's cross-border import retail e-commerce includes business-to-consumer (B2C) and consumer-to-consumer (C2C) modes. According to Moda data, the proportion of B2C and C2C is 2.1% and 97.9% in 2013, but this proportion changed to 64.4% and 35.6% in 2017 [2]. So B2C has surpassed C2C as the most important cross-border e-commerce import retail mode in China, and the cross-border e-commerce we discuss here is based on this B2C mode.

In 2013, President Xi Jinping proposed to build the "The Belt and Road Initiative (B&R)". The E-Commerce Connectivity Index (ECI) reflects the close connection between China and B&R countries, the higher ECI import index the country gets, the more Chinese consumers buy products from that country. According to AliResearch, Thailand's ECI import index is 6.9, which is ranking first among all B&R countries [3]. It indicates that Chinese consumers have the highest purchasing preference to buy Thai products among B&R countries.

II. COSMETICS SALES SITUATION ON CHINA CBEC PLATFORM

China has become world's second largest cosmetics consumption country since 2013. As the iiMedia Research showed that 47.4% Chinese consumers buy cosmetics from a comprehensive e-commerce platform [4], CBEC has become the mainstream way for Chinese consumers to buy cosmetics. Yi Guan Analysis reported that Tmall Global has ranked first with a market share of 32.3% in the first quarter of 2019 [5]. According to Aliresearch, the most popular CBEC product category on Tmalll Global was cosmetics which reaching 32% in 2018 [6].

Aliresearch also reported Thailand has become the highest-sales ranking country among B&R countries with a year-on-year grow increase of 125%. Especially, Thai cosmetics have become the most popular category for Chinese consumers, accounting for 56.1% of total sales in

2018 [7]. However, it is worth noting that CBNData showed the largest three cosmetics importing countries were South Korea, Japan and France from 2017 to 2018 on Tmall Global, while Thai cosmetics sales was ranked outside the top ten [8], which was far behind from the top three countries.

III. MEDIA RESEARCH ALSO REPORTED THE MAIN FACTOR AFFECTING CHINESE CONSUMERS' PURCHASING DECISION OF CROSS-BORDER COSMETICS IS WORD OF MOUTH, ACCOUNTING FOR 57.3% RESPECTIVELY, WHILE OTHER FACTORS WERE LESS THAN THIS PROPORTION [4]

From industry data, on one hand, we can see that even on the same CBEC platform, Chinese consumer's purchasing preference of cosmetics is far different. They prefer to buy cosmetics from South Korea, Japan, and France rather than Thailand. So, it is significant to analyze Chinese consumer concerns of cosmetics. On the other hand, electronic word-of-mouth (eWOM) also has a huge impact on Chinese consumer purchasing behavior, and it is expressed by massive online reviews through CBEC platform, which needs to be understood well.

TABLE I: COSMETICS SALES SITUATION ON CHINA CBEC PLATFORM

Cosmetics purchase channels	Comprehensive e-commerce platform (47.4%)	Cosmetics store (42.2%)	Department store counter (39.7%)
Factors influencing purchasing decision	Word of mouth (57.3%)	Brand (56.5%)	Function (47.8.8%)
Top3 CBEC platform in Q1 2019	Tmall Global (32.3%)	Kaola (24.8%)	JD Worldwide (11.6%)
Top3 popular category	Cosmetics (32%)	Maternal and child products (19%)	Health products (17%)
Top3 cosmetics importing countries on Tmall Global (2018)	Korea	Japan	France
Thai popular category	Cosmetics (56.1%)	Health products (16%)	Household products (15.9%)

A. Problem Identification of CBEC Consumer Concerns

This part aims to identify research problems from existing literature, and how these problems have been addressed.

The table above shows what previous researches have done. In Europe context, Gomez and Martens have examined the drivers factor for CBEC from a consumer survey, they used gravity model to analyze and found that drivers will influence consumer purchase including language-related trade cost, parcel delivery and online payment [9]. Cardona and Duch have conducted a questionnaire to show consumer perception in the digital single market by using regression analysis and cross-border bilateral online trade matrix, then they found products variety is the most critical factor rather than price [10].

When turning back eyes to Asian countries, there are also some researchers study of CEBE consumer concerns. In Korea, Han and Lee explored the consumer attitudes by questionnaire, they applied an OLS regression-based named partial least squares to evaluate model and found E-service quality had influence of cross-border online shopping [11]. Lin and Lee from Taipei used partial least square method for testing the proposed model based on online survey, and they

found that service justice was the most important factor affecting people of CBEC purchase [12]. In China, Hsiao *et al.* examined that logistics service influence people to engage in CBEC activity by Kansei engineering and partial least squares [13]. Li and Yao conducted a comparative study among nine B&R countries by using LDA model and sentiment analysis, and they found six major factors from massive online reviews, then examined the consumer's risk perception from negative views [14]. Mou *et al.* also concerned consumers feedback and explored 35 primary topics by using LDA model in CBEC context, and they explored from two sides, sellers concerned more about regarded commission, product audit and communication while buyers concerned more about return and fund, product tracking and product description [15].

TABLE II: CBEC ONLINE CONSUMER CONCERNS

Author	Country	Topic	Method
Gomez &Martens (2014)	EU	The drivers and impediments for cross-border e-commerce in the EU	1.Descriptive statistics derived from a consumer survey 2. Analytical tool- gravity model
Cardona &Duch (2015)	EU	Consumer perception of (cross-border) e-commerce in the EU digital single market	1.Questionnaire for data collection 2.Regression analysis to estimate the relative importance of different drivers 3. A cross-border bilateral online trade matrix and to estimate a gravity trade model
Hsiao et al (2017)	China	Logistics service design for cross-border E-commerce using Kansei engineering with text mining-based online content analysis	1.Online content text-mining 2.Kansei engineering (KE) for designing the service elements 3. Partial Least Squares (PLS) for analyzing the relationships between the feelings of customers and elements
Han &Lee (2018)	Korea	Exploring consumer attitudes and purchasing intentions of cross-border online shopping in Korea	1.Questionnaire for data collection 2.Partial least squares-an OLS regression-based estimation method for evaluating model
Lin &Lee (2018)	Taipei	Dysfunctional customer behavior in cross-border e-commerce: a justice-affect-behavior model	1.Online survey for data collection 2.The partial least square method for testing the proposed model.
Li &Yao (2019)	China	Consumer's risk perception on the Belt and Road countries: evidence from the cross-border e-commerce	1.LDA model for risk factors extraction 2.Sentiment analysis for sentiment score
Mou et al. (2019)	China	Understanding the topics of export cross-border e-commerce consumers feedback: an LDA approach	LDA model for exploring the topics

From above summary, we can find valuable research room for this research. Firstly, more and more researchers study the consumer concerns and sentiment preference based on massive online reviews, but most them only study within one country. Secondly, some researchers also concern about CBEC context, but they explore more on cross-border export e-commerce rather than import, there is a limited research on China cross-border import ecommerce. Thirdly, when to explore Chinese consumer concerns of comparative study between Thailand and other foreign countries, and there is no one compare them in cosmetics industry. Fourthly, the main data collection method existing is questionnaire rather than big data technology such as LDA and sentiment analysis in CBEC context.

Based on industry data and literature review, there are two current specific research problems are identified:
1) Problem 1: For Thai, Japanese, Korean and French cosmetics, what are the differences in Chinese consumer concerns?
2) Problem 2: For Thai, Japanese, Korean and French cosmetics, what are the sentiment preferences in Chinese consumer concerns?

According to two research problems, there are two research objectives:
1) Objective 1: Identify Chinese consumer concern factors and importance of Thai, Japanese, Korean and French cosmetics,
2) Objective 2: Identify the sentiment preference in

Chinese consumer concern factors of Thai, Japanese, Korean and French cosmetics.

This part justified some key points for comparative study: First of all, we justified Tmall Global is the largest CBEC platform in China, and cosmetics is the most popular category. Secondly, we justified four comparative countries were Thailand, Japan, Korea and France. Thirdly, we found that consumer concerns were discussed less in CBEC context rather than domestic e-commerce. Fourthly, some big data technology can be applied to analyze consumer concerns.

IV. LITERATURE REVIEW

A. Electronic Word of Mouth(eWOM)

Litvin and Pan explained eWOM is "all informal communications directed at consumers through Internet based technology related to the usage or characteristics of particular good and services, or the sellers" [16], and they studied in hospitality and tourism management.

Trusov and Pauwels did a research to study effects of Internet social networking site, and they found that eWOM is more effective than traditional marketing [17]. Sotiriadis and Zyl studied the eWOM and online reviews in tourism services and found that customers influenced by online customer reviews [18].

Positive WOM show the satisfying purchase experiences and negative reviews show complaints. Ghose studied the influences between customer reviews and sales by analyzing the sentiments and found that many embedded properties of reviews matter in impacting sales [19].

In summarization, eWOM has a critical impact in different industries, here it also can be applied in CBEC context, and their positive and negative reviews can be analyzed for further development for some parties.

B. Latent Dirichlet Allocation and Sentiment Analysis

Latent Dirichlet Allocation (LDA) is a statistical model used to discover the underlying abstract topics in a series of documents or text data. Based on "bag of words', every document has some latent topics and each topic is a multinomial distribution over words. It can see each topic's proportion in each document and the word distribution in each topic [20].

Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [21]. This method is widely used to voice of the customer materials such as reviews and survey responses, online and social media.

Massive consumer reviews show their concerns. However, most of researchers conduct survey to collect data based on questionnaire, which is not always effective under every condition, especially during the cross-border e-commerce purchasing environment. Firstly, the accuracy is strongly influenced by the data collected because of place and time. It is quite hard to touch target consumer with large scale in short time. Secondly, from the researcher perspective, the questionnaire designed is influenced by researcher themselves, and it is hard to say whether truly reflect responder aspirations. Thirdly, from consumer perspective, it is a common problem in any questionnaire survey that some consumers are not willing to fill it or do not show their real emotion.

Recently, some researchers have done similar research depends on big data technology and try to find consumer real concerns from online reviews. The table below shows the study topic and comprehensive methodology explored.

TABLE III: LITERATURE REVIEW OF METHODOLOGY

Author	Title	Main Method
Chen and Xu (2016)	The determinants of online customer ratings: a combined domain ontology and topic text analytics approach.	1.Product ontology for describing product aspect 2.Latent Dirichlet Allocation (LDA) Model for topic extraction 3.Aspect-orient sentiment analysis for sentiment polarities and score 4.Multiple regression model for ratings
Al-Obeidat (2017)	The opinion management framework: Identifying and addressing customer concerns extracted from online product reviews.	1.LDA for topic mapping 2.Sentiment analysis for frequency of positive and negative words 3.Cost/benefit analysis
Wang (2018)	Topic analysis of online reviews for two competitive products using Latent Dirichlet Allocation.	LDA for topic extraction and analysis
Li and Yao (2019)	Consumer's risk perception on the Belt and Road countries: evidence from the cross-border e-commerce	1.LDA model for risk factor extraction 2.Sentiment analysis (NB) for sentiment score
Mou (2019)	Understanding the topics of export cross-border e-commerce consumers feedback: an LDA approach	LDA model for exploring the topics

In 2016, Chen and Xu aimed to study the determinants of online customer rating which was using innovative aspect extraction method. They provided product ontology and combined the results of the topic modeling method LDA to get 8 aspects, then multi-aspect-oriented sentiment analysis was used to measure the sentiment polarities. Multiple regression analysis was also used to link the customer reviews and numerical rating [22].

In 2017, Al-Obeidat proposed an opinion management framework to identify customer concerns extracted from online product reviews, they also introduced LDA for topic mapping and sentiment analysis for the frequency of positive and negative words and tried to use cost/benefit analysis to provide advice for business [23].

In 2018, Wang has conducted topic analysis of online reviews for two competitive products by using LDA, and their different topics and relative importance are identified by comparative study [24].

In 2019, Li and Yao have conducted a research by using LDA and sentiment analysis. Six topics were extracted, and consumer perceptions were calculated by negative reviews. They did the comparative analysis among nine B&R countries [14]. Mou applied the LDA for exploring the topics of export CBEC from both sellers and buyer's perspective, and they found 35 topics from two sides [15].

In summarization, there are some main points. Firstly, comparing with questionnaire, researchers began to analyze consumer concerns by using big data mining methods including LDA and sentiment analysis. Secondly, these two methods are used single or together based on different researcher objectives, but there is a limited case to combine them in CBEC context, and no one apply them in CBEC cosmetics industry before.

V. METHODOLOGY

The objectives of this research are to identify Chinese

consumer concern factors and their sentiment preference of Thai, Japanese, Korean and French cosmetics. In this section, the methodology designed will elaborate from sample products selection and text-mining based conceptual framework. On one hand, sample selection determines what kind of data we input. On the other hand, this conceptual framework determines how to realize two objectives of extracting Chinese consumer concern factors and identify their positive and negative attitude exactly.

A. Sample Product Selection

The textual online reviews used in this research are collected from Tmall Global, which is one of the largest CBEC platform in China for selling various of importing products. To identify Chinese consumer concerns among Thai, Japanese, Korean and French cosmetics, best seller products of each country are selected as the representative products to get the online reviews.

Here researcher will apply keywords searching of each country's name in Chinese, the target sample products will be selected by monthly sales volume in cosmetics category. It is worth noting that Tmall Global has anti-reptile protection, in order to have enough online reviews for final analysis, top5 sales volume cosmetics will selected in each country. The main steps are described as below:

1) Step 1: Search keywords with each country's name in Chinese on Tmall Global official website.
2) Step 2: Choose product display method is "sales volume".
3) Step3: Select top5 products with highest monthly sales volume in cosmetics category.

B. Text-mining Based Conceptual Framework

As previous literature review show that massive textual online reviews can reflect consumer concerns after their purchase experience, in turn, their reviews also affect other consumers purchase willing. This part aims to propose a text-mining based conceptual framework which including research objectives, methods used and detail process. In order to extract the consumer concern factors and identify their sentiment polarity of these four countries, researcher here combines LDA and sentiment analysis to reach research goal (see Fig. 1).
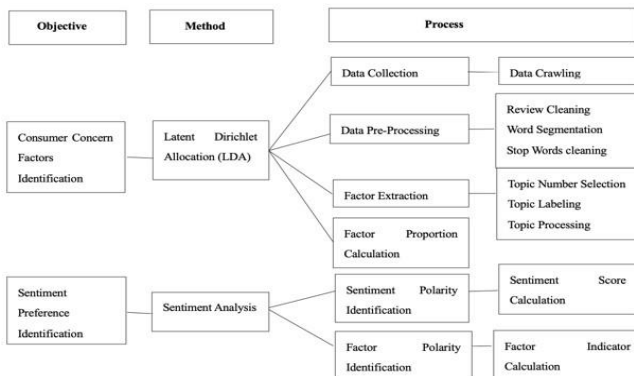

Fig. 1. Text-mining based conceptual framework.

First of all, LDA is used to identify the consumer concern factors and obtain their proportion. Some important steps need will be described as data collection, data pre-processing, factor extraction and factor proportion

calculation.

1) Data collection. Target sample products are selected after the sample selection step, then consumer online reviews are collected by crawling data from Tmall Global. This is the raw data without any processing step.
2) Data pre-processing. This step aims to clean unnecessary data and improve the accuracy of the results. Firstly, it is necessary to delete useless reviews including incomplete, repeated or irrelevant words [22]. Secondly, suitable word segmentation tool needs to be chosen. Based on the fact that the consumers from Tmall Global are Chinese, online reviews are written in Chinese as well, so a Chinese segmentation tool named Jieba will be used. Thirdly, it also needs to remove some stop words such as 'then" at" it', and then three widely used Chinese stop words list are selected which are from Harbin Engineering University, Machine Intelligence laboratory of Sichuan University and Baidu [14].
3) Factor extraction. The consumer concern factors are extracted by using LDA. Firstly, Gibbs sampling and perplexity are used when making topic selection. When the perplexity tends to converge, that means the topic number tends to suit, and the smaller number of topics is better according to the perplexity scores [25], but it needs to make a balance when choosing topic number for better topic explanation. Secondly, topic labeling is a critical step by conducting manually according to the logical connection between their top words and the relative weights. Thirdly, we also consider doing topic processing by deleting or combining topics to identify meaningful topics.
4) Factor proportion calculation. This step needs to be done by calculating the proportion of reviews classified into each concern factor in each country. That means total reviews will be classify to different factors, the number of reviews in each factor occupy the total reviews' proportion represent each factor importance. The larger proportion value number means the higher importance level of this concern factor, the lower proportion value number means the lower importance level of this concern factor.

Secondly, sentiment analysis method is applied for identifying Chinese consumer concern sentiment preference by calculating the sentiment score of each factor. Each review has its sentiment score, and the value is between 0 to 1, it indicates the sentiment polarity. From 0 to 0.5 represents negative, then from 0.5 to 1 represents positive, and 0.5 represents mediate. Thus, the positive concern indicator and negative concern indicator can be identified by calculating the positive and negative reviews of each factor and weighting the probability of indicator.

VI. RESULTS

In this section, we will show the selected sample products and do some analysis from their sales volume and reviews, then the proposed text-mining based conceptual framework output will be discussed.

A. Selected Sample Products and Analysis

In last section, we explained the target sample products selected by keywords searching with top five highest sales volume in cosmetics category. Specifically, when we search with '泰国 (Thailand)', '法国 (France)', '日本 (Japan)' and '韩国 (Korea)' on Tmall Global platform, a lot of products will be displayed on the website pages, and we select "sales volume" as product ranking method manually, and then four selected countries and their top5 sample products brand and name, monthly sales volume and customer reviews are summarized in table below.

As the results show, 25 sample products are selected. The total highest monthly sales volume country is Japan with 577000 pieces, and then is Korea, France and Thailand with 514000, 44200 and 165000 pieces. It can interpret that Chinese consumer prefer to choose cosmetics more from Japan, Korea and France rather than Thailand.

Additionally, from the online reviews' column, we can see Thai cosmetics has highest total consumer reviews with 626497 reviews, and then is Japan, Korea and France with 403271, 347925, 267187 reviews. So, we can see the relationship between consumer reviews and sales volume. The rate in the table shows every review contribute to the sales volume (Rate=total sales volume /total reviews). The ratio of Japan, South Korea and France is 1.43, 1.48 and 1.65, that means there is a significant positive correlation between product sales volume and reviews in these countries, while the rate of Thailand is 0.26 which means the sales volume is negative correlated with reviews.

TABLE IV: SAMPLE PRODUCTS INFORMATION OF 4 COUNTRIES

Country	Product	Monthly Sales Volume (piece)	Online Reviews	Ratio
Thailand	Mistine Little Yellow Hat Sunscreen	73000	370747	
	Mistine Little Blue Hat Body Sunscreen	29000	141488	
	Mistine Blue Shield Liquid Foundation	27000	14345	
	Mistine Liquid Eyeliner Pen	19000	3898	
	Beauty Buffet milk facial cleanser foam	17000	96019	
	Total	165000	626497	0.26
France	French Bioderma Cleansing Water	180000	107992	
	French La Roche-Posay Lotion Spray	80000	12649	
	Lancome large pink water	77000	24670	
	Avene Spray	69000	64099	
	La Roche-Posay B5 Repair Cream	36000	57777	
	Total	442000	267187	1.65
Japan	Freeplus acid facial cleanser foam 100g	320000	124496	
	Curel facial cleanser	71000	60390	
	Curel moisturizing lotion	69000	63629	
	Shiseido Fino Hair Mask Conditioner	60000	134893	
	Ettusais mascara base	57000	19863	
	Total	577000	403271	1.43
Korea	innisfree mineral loose powder	130000	46448	
	Amore foam hair dye	120000	183198	
	3CE Lip Glaze	100000	3086	
	AHC face sunscreen	88000	106863	
	Romand Lip Glaze	76000	8330	
	Total	514000	347925	1.48

It can interpret that Chinese consumer has strong concerns and express their sentiment of products in each country, especially Thai products. As previous literature review, positive reviews increase product sales while negative reviews decrease products sales. Therefore, it indicates that there are more negative reviews to express consumer sentiment preference inside Thai products comparing with other three countries.

Since Tmall Global has anti-reptile protection, the crawlable review number of each product is 1980, so each country can crawl 9900 reviews. Some review examples of sample products crawled are listed in the following table.

Generally speaking, the content of these reviews expresses the factors that consumers care about and their sentiment preference towards these factors. We take reviews from Thailand as example, "Good, non-greasy, with moisturizing effect and also whitening" show consumer is satisfied with function, while "It will experience pilling, maybe because I used too much, the logistics was not good" show consumer complain about the feeling of use and logistics. These reviews will be used as input raw data and analyzed in detail for further research.

TABLE V: REVIEW EXAMPLE OF FOUR COUNTRIES

Country	Review Example
Thailand	Good, non-greasy, with moisturizing effect and also whitening（好用不油腻，带保湿效果,还美白）
	It will experience pilling, maybe because I used too much, logistics was not good（会搓泥，可能是我用量过大，快递不给力）
France	Remove makeup clean and refreshing（卸妆干净，清爽）
	I chose the mild, but sent me oil control, not careful at all（我选的温和的，给我寄的控油的，一点不细心）
Japan	The price is affordable, cheap during promotion（价格实惠，活动很便宜）
	Barcode has problem, I'm a bit worried about whether it is genuine（划码的，有一点担心是不是正品）
Korea	Purchased many times, the oil control effect is great（多次购买，控油效果棒棒哒）
	Not lasting, not moisturizing（不持久，不保湿）

B. Text-Mining Based Conceptual Framework Output

In this section, the text-mining based conceptual framework output will be discussed. The flowchart below describes the results after implementing the whole process. In order to identify the objective of consumer concerns and their sentiment preference, online consumer reviews in each country are input for operation and some results are obtained (see Fig. 2).
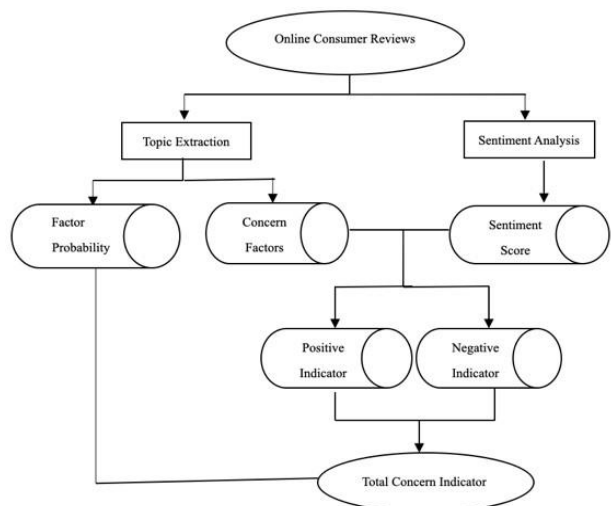

Fig. 2. Flowchart of output.

First of all, topic extraction will be done by LDA method. There are two elements obtained which are concern factors and their probability. The topic extraction results can be discussed from 3 perspectives: topics extracted, related importance and topic difference.

Specifically speaking, topics extracted can show how many concern factors affecting consumer purchase of

cosmetics and what they are in each country. Related importance can show each concern factor importance occupation in each country, then we will do a ranking of these factors to give indications both for consumers and suppliers. Topic difference can be done by comparing the concern factors among Thailand, Japan, South Korea, and some important words that consumer concern also can be showed by word cloud, then some suggestions can be given for Thai cosmetics comparing other three countries' cosmetics.

Secondly, sentiment analysis is used to identify consumer sentiment preference. The results of this step will show the sentiment polarity. Specifically speaking, score value of each review in each country can be calculate and all of them can be identified as positive or negative review.

Thirdly, concern factors and sentiment score will be combined to get concern factors' sentiment indicator and their indicator difference. As the sentiment value of each review has been identified in the last step, this step is to classify the positive and negative reviews in each of these factors in each country, and we can see the respective proportion weight of the positive and negative reviews to determine whether this factor is positive or negative concern factor, then the comparison among four countries will be done to indicate each country's advantage and disadvantage. This aims to give some suggestions for Thai cosmetics suppliers to improve their poor weaknesses by analyzing another three countries strengths, at the same time Thai cosmetics suppliers can maintain their strengths and find their competitiveness by analyzing another three countries' weaknesses.

VII. CONCLUSION

Cross-border e-commerce has an important impact to promote economic integration and trade globalization, which is developing quickly along B&R countries.

This research contributes in three parts. First of all, we identified cosmetics sales situation on China largest CBEC platform, two important research problems and objectives are obtained, which is to examine the Chinese consumer concerns and sentiment preference. Secondly, five sample products and 9900 consumer reviews are selected from Thailand, Japan, Korea and France, which can be input data for the further study. Thirdly, a text-mining based conceptual framework is introduced in the CBEC cosmetics context including LDA method and sentiment analysis. LDA is helpful for extracting consumer concern factors and their proportion, and sentiment analysis can show the sentiment polarity of individual sentiment score of each factor.

Who can benefit from this research? When we look back from the cross-border e-commerce process, there are some main participants including products supplier, platform operator, logistics provider, and so on. They want to identify consumer concerns factor and their positive and negative attitude on CBEC cosmetics category. This research especially can benefit participants along B&R countries by comparing study. It can indicate the customer concerns gap between country from B&R road and popular cosmetics country.

This research also has limitation. Here we don't consider how concern factors and preferences change when consumer reviews change with time.

In the future work, researcher will implement proposed text-mining conceptual framework and import data crawled from Tmall Global to further detail analysis. We are looking forward to giving some policy making recommendations in the CBEC cosmetics context after analysis.

Conflict of INTEREST

The authors declare no conflict of interest.

Author CONTRIBUTIONS

ACKNOWLEDGMENT

REFERENCES

[1] B. Hu and Q. Q. Luo, "Cross-border e-commerce mode based on internet +," in *Proc. IOP Conference. Series: Materials Science and Engineer*, 2018.
[2] *Analysis of Cross-Border Import e-Commerce*, Moda Data, 2018.
[3] *eWTP Helps the Construction of the Belt and Road Initiative*, AliResearch, 2017.
[4] 2019 China's Cosmetics e-Commerce Industry Monitoring Analysis and Comprehensive Survey Report on Consumer Behavior, iiMedia Research, 2019.
[5] 2019 Q3 Cross-border E-commerce Market Share Report, Yi Guan Analysis, 2019.
[6] China Import Consumer Market Research Report, AliResearch, 2019.
[7] Building the 21st Century Digital Silk Road: The Practice of Alibaba Economy, AliResearch, 2018.
[8] Cross-Border Import Consumption Insights Report, CBNData, 2018.
[9] E. Gomez-Herrera, B. Martens, and G. Turlea, "The drivers and impediments for cross border e-commerce in the EU," *Information Economics and Policy*, vol. 28, pp. 83–96, 2014.
[10] *Consumer Perceptions of (Cross-border) e-Commerce in the EU Digital Single Market*, Joint Research Centre, Working paper, European Commission, 2018.
[11] B. Han, M. Kim, and J. Lee, "Exploring consumer attitudes and purchasing intentions of cross-border online shopping in Korea," *Journal of Korea Trade*, vol. 22, no. 2, pp. 86–104, 2018.
[12] A. J. Lin, E. Y. Li, and S.-Y. Lee, "Dysfunctional customer behavior in cross-border e-commerce: A justice-affect-behavior model," *Journal of Electronic Commerce Research*, vol. 19, pp. 36–54, 2018.
[13] Y.-H. Hsiao, M.-C. Chen, and W.-C. Liao, "Logistics service design for cross-border E-commerce using Kansei engineering with text mining-based online content analysis," *Telematics and Informatics*, vol. 34, pp. 284–302, 2017.
[14] J. P. Li, Y. H. Yao, Y. J. Xu, J. Y. Li, L. Wei and X. Q. Zhu, "Consumer's risk perception on the Belt and Road countries: evidence from the cross-border e-commerce," *Electronic Commerce Research*, vol. 19, pp. 823–840, 2019.
[15] J. Mou, G. Ren, C. X. Qin and K. Kurcz, "Understanding the topics of export cross-border e-commerce feedback: An LDA approach," *Electronic Commerce Research*, vol. 19, pp. 749–777, 2019.
[16] S. W. Litvin, R. E. Goldsmith, and B. Pan, "Electronic word-of-mouth in hospitality and tourism management," *Tourism Management*, vol. 29, pp. 458-468, 2008.
[17] M. Trusov, E. R. Bucklin, and K. Pauwels, "Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site," *Journal of Marketing*, 2009.
[18] M. D. Sotiriadis and C. van Zyl, "Electronic word-of-mouth and online reviews in tourism services: The use of twitter by tourists," *Electronic Commerce Research*, vol. 13, pp. 103–124, 2013.

[19] A. Ghose and P. G. Ipeirotis, "Designing novel review ranking system: predicting the usefulness and impact of reviews," in *Proc. the Ninth International Conference on Electronic Commerce*, 2011, vol. 23, pp. 1498-1515.

[20] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, 2003.

[21] L. Bing, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, pp. 1-167, 2012.

[22] R. Chen and W. Xu, "The determinants of online customer ratings: A combined domain ontology and topic text analytics approach," *Electronic Commerce Research*, vol. 17, no. 1, pp. 31–50, 2016.

[23] F. Al-Obeidat, B. Spencer, and E .Kafeza, "The opinion management framework: Identifying and addressing customer concerns extracted from online product reviews," *Electronic Commerce Research and Applications*, vol. 27, pp. 52–64, 2018.

[24] W. Wang, Y. Feng, and W. Dai, "Topic analysis of online reviews for two competitive products using Latent Dirichlet allocation," *Electronic Commerce Research and Applications*, vol. 29, pp. 142–156, 2019.

[25] L. Wei, G. Li, X. Zhu, X. Sun, and J. Li, "Developing a hierarchical system for energy corporate risk factors based on textual risk disclosures," *Energy Economics*, vol. 80, pp. 452–460, 2019.

Yuting Feng was born on January 28th, 1992. She is from Anhui China. She got two bachelor's degrees in accounting and law from Zhengzhou University of Light Industry, China. She received master's degree in knowledge management from Chiang Mai University, Thailand. Now she is a PhD student in big data and financial technology at Chiang Mai University, Thailand. Her research areas include knowledge management, cross-border e-commerce, big data and financial technology.

She is working as a research assistant for the project of Thailand-China cross-border e-commerce at International College of Digital Innovation, Chiang Mai University, Thailand. She also published paper titled "Competency requirements of managerial level logistician in Northern Thailand" in *International Journal of Management and Applied Science*, vol. 4, issue 7, pp. 20-24, 2018.

Miss Yuting Feng has been awarded full scholarship for Ph.D. at International College of Digital Innovation, Chiang Mai University in 2019.

Nopasit Chakpitak got his bachelor's degree in electrical engineering from Chiang Mai University, Thailand in 1986. He received his Ph.D. in electronic and electrical engineering from University of Strathclyde, Scotland in 2002.

He is a researcher in artificial intelligence, knowledge engineering and management. His working experiences are listed as following: a) assistant to the president in International Academic Affairs during 2012-2014; b) dean in the College of Arts, Media and Technology, Chiang Mai University, Thailand during 2004-2011; c) NECTEC IT program coordinator for NSTDA (National Science and Technology Development Agency) Northern Network during 1998-200; d) Lecturer in Chiang Mai University from 1993 until now; e) power plant engineer, Electricity Generating Authority of Thailand during 1987 to 1992; f) project engineer in Mass Communication Organization of Thailand during 1986 to 1987.

Asst. Prof. Dr. Nopasit Chakpitak now works as the dean of International College of Digital Innovation, Chiang Mai University, Thailand. He is also the founder dean of the College of Arts, Media and Technology, Chiang Mai University. Now he is an assistant professor in knowledge management and software engineering. He provides consulting services in knowledge management implementation for government agencies, state and private enterprises by using knowledge engineering methodology. He also involves in several European funded projects in ICT and its applications. His current works are on smart city including smart transportation, smart tourism, smart education, smart energy, etc.

Piang-or Loahavilai was born in Bangkok, Thailand. Her first degree was in political science. She was a Korean government scholarship fellow and graduated with the M.B.A from Seoul National University in Seoul, Republic of Korea. She continued her Ph.D. in knowledge management at Chiang Mai University in Thailand. Dr. Piang- or Loahavilai is the president of Global Korean Scholarship (GKS) Alumni, Thailand Chapter and a GKS publicity envoy appointed by the Ministry of Education of the Republic of Korea. In 2019, she received a minister award from the Ministry of Tourism and Culture of the Republic of Korea for her 500 pieces of publication of Korean related articles and books for 10 years. Currently, she is also a consultant of the Minister's office of the Ministry of Transportation of Thailand.

Tirapot Chandarasupsang was granted with the Ph.D. in electronics and electrical engineering from the University of Strathclyde, UK in 2005. His research interests include power system economics and electricity market designs, asset management, knowledge management and engineering, and modelling and simulation techniques.

He is a Senior lecture at International College of Digital Innovation, Chiang Mai university, Thailand, and he also provide consulting services for both government and private organizations.

Lei Mu got his bachelor of engineer in information engineering from the University of Electronic Science and Technology of China, P. R. China in 2003. He received the master of engineer in IT engineering management from the University of Electronic Science and Technology of China, P. R. China in 2011. He was granted Ph.D. degree in computer science from the University of Grenoble Alpes, France with full scholarship supported by the Erasmus Mundus Action 2: Sustainable e-Tourism.

He was a director of Modern Language Educational Technology Center of the College of Foreign Languages and Cultures, Chengdu University, China and now work as the deputy dean of the College of Foreign Languages and Cultures, Chengdu University, China. His research interests include e-tourism, mobile user privacy and e-commerce.

Dr. Mu Lei is a committee member of international conference named SKIMA, as well as invited expert in a Chinese Journal, Education and Teaching Research.