Research on Demand Mining Method for Short Life Cycle Experiential Products Based on Structural Topic Model and Experience Value

Zhongjun Tang and Xinhao Zhou
School of Economics and Management, Beijing University of Technology, Beijing 100124, China

Keywords: Short Life Cycle Experience Products, Demand Mining, Structural Topic Model, Experience Value.

Abstract: With the development of social economy, short life cycle experiential products occupy an increasingly important position in the market. The demand for short life cycle experiential products through big data methods is of great significance to product improvement and innovation. This paper proposes a short-life cycle experience product demand mining method based on structural topic model and experience value. In order to meet the short-term characteristics of short life cycle experiential products, collect user comment data on the website, and adopt the structural topic model (STM) method, the user comment rating is used as the covariate in STM model, extract the customer demand topic words and their corresponding emotional tendencies and visualizing. The demand topics excavated using the STM method are divided into five categories based on the experience value theory, so that the excavated short life cycle experiential product demands are experiential. This paper takes a movie as an example to verify the effectiveness of the proposed method. The method is effective and more accurate than traditional methods, which provides guidance for enterprises to tap customer needs and product innovation.

1 INTRODUCTION

With the continuous change of consumer demand for products and the rapid development of product production technology, the update speed of the products is accelerated, the innovation cycle is shortened, and the short life cycle experience products occupy an increasingly important position in the market. Short life cycle experiential products have the characteristics of short-term and experiential. The short-term is mainly reflected in the rapid renewal of products, the rapid decline of value, and the strong substitution effect of competitive products. The experiential is mainly reflected in that consumers can only evaluate the product quality after consumption, and the subjective participation is strong. Movies and mobile games are typical short life cycle experience products (TANG, CUI, TANG, ZHU 2019). It is of great significance to study the method of demand mining for short life cycle experiential products, and to discover and analyze customer demand for real enterprises, so as to improve and innovate products.

Traditional market research methods such as structured questionnaire or semi-structured questionnaire are mainly used in existing research on customer demand mining. This kind of research method requires high rationality in questionnaire design and takes a long time to design questionnaire, which is not suitable for short life cycle experience products. With the development of the Internet and online comment platform, customers will choose the network platform to make subjective comments after consumption. Online comments are directly from customers, which have the characteristics of spontaneity, authenticity, and have become a research hotspot of researchers. Using text mining method to mine customer demands from online reviews can help manufacturers innovate products quickly. Combined with customer experience value, this paper constructs demand classification according to the experiential characteristics of short life cycle experiential products. In the same movie, when the same keyword appears as a high-frequency word in both positive and negative comments, it is difficult for researchers to judge its true emotional tendency. Therefore, on the basis of the existing online review demand mining methods, we also demand to solve the problem that the frequency of keywords in positive comments is equal to that in negative comments, which makes it impossible to judge the real tendency of a keyword in online

Tang, Z. and Zhou, X.
Research on Demand Mining Method for Short Life Cycle Experiential Products Based on Structural Topic Model and Experience Value.
DOI: 10.5220/0011754200003607
In *Proceedings of the 1st International Conference on Public Management, Digital Economy and Internet Technology (ICPDI 2022)*, pages 665-673
ISBN: 978-989-758-620-0

comments.

In summary, this article proposes a short-life cycle experience product demand mining method. This method uses online reviews as the data source, replaces traditional text mining methods with structural topic models to improve the accuracy of demand mining, and compares the results with traditional text mining methods to illustrate the importance of using of structural topic models. Then combine the experiential value theory with the unearthed demand topics. Take a movie as a case to verify the effectiveness of the method proposed in this article.

2 RELATED WORKS

2.1 Topic Model and Online Comment Demand Mining Method

The existing online comment data demand mining methods mainly focus on text mining and sentiment analysis, and mainly use content analysis and machine learning methods to conduct topic mining on online comments (Sangwan, Malik, Sunesh, et al. 2017), sentiment polarity judgment (Usama, Ahmad, Song, et al. 2020), and sentiment intensity calculations. The common method of topic mining is the Latent Dirichlet Allocation (LDA)model. Researchers can mine the focus of users through LDA, but the emotions implied in the comment cannot be judged. The sentiment polarity judgment and sentiment intensity calculation are usually based on the sentiment dictionary. But the method has high requirements for the construction of sentiment dictionary.

On the basis of the above, the researchers put forward a method to combine the LDA model with sentiment analysis to simultaneously obtain the topic and the emotion corresponding to the topic from the comments. Lin et al. (Lin, He 2009) added an additional sentiment layer to the LDA framework and established a joint sentiment/topic model (JST), which links sentiment polarity, documents, topics, and words. Jo et al. (Jo, Oh 2011) proposed the Aspect and Sentiment Unification Model (ASUM) to extract feature-sentiment pairs from documents. Dong et al. (Dong, Ji, Zhang, et al. 2018) proposed an unsupervised topic-sentiment joint probabilistic model (UTSJ) on the basis of JST. After improving the topic model, Rao et al. (Rao, Li, Mao, et al. 2014) and Feng et al. (Feng, Rao, Xie, et al. 2019) proposed the Sentiment Latent Topic Model (SLTM) and the Multi-label Supervised Topic Model (MSTM), the accuracy of emotion calculation in the topic-emotion model is improved by adding different types of supervision tags to different generation stages of the model.

The above research mainly focuses on the improvement of the accuracy and stability of the emotion polarity extraction and calculation in the topic-emotional model, and does not consider the frequency difference of keywords in negative comments and positive comments.

2.2 Structural Topic Model (STM)

Roberts et al. (Roberts, Stewart, Tingley, et al. 2014) proposed structural topic model (STM) in 2014, which effectively solved the problem of keyword frequency of positive and negative reviews. Structural topic model allows researchers to input some features of reviews into the model as covariate parameters in advance to explore the relationship between topics and review features. Taking the positive and negative tendency of reviews as covariates can overcome the frequency problem of the above keywords, mine customers' real satisfaction and dissatisfaction demands, and improve the accuracy of demand mining method. As shown in Figure 1, the topic popularity covariate X affects the document-topic probability generation, and the topic content covariate Y affects the topic-word probability generation.

STM has been used in the text analysis of news reports (Chandelier, Steuckardt, Mathevet, et al.,

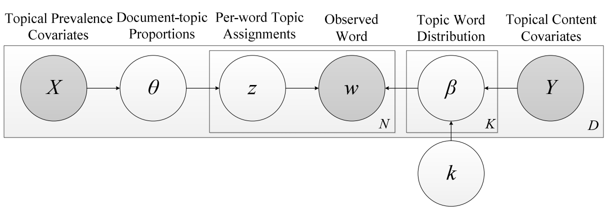

Figure 1: Schematic diagram of structural topic model.

2018), political discourse (Curry, Fix 2019) and other fields, using time as a covariate to analyze the trend of thematic changes in news reports. Scholars such as Tvinnereim (Tvinnereim, Fløttum 2015) used STM to extract topics from semi-structured texts such as question-and-answer texts (Lester, Kessler, Modisett, et al. 2019) and aviation reports (Kuhn 2018), in order to study the user characteristics obtained from semi-structured texts and the relevance of texts. In view of the above-mentioned differences between STM and LDA, the use of STM to mine online user reviews of experiential products has the following two advantages: 1. STM enables researchers to analyze document-level covariates (such as whether the review is positive or negative, the time of the review, The gender of the reviewer, etc.) introduce topic popularity covariates to explore the impact of the covariate on the document-topic probability, and at the same time, explore the changes in the document-topic probability with the change of the covariate, and consider the heterogeneity of positive and negative reviews. 2. STM enables researchers to introduce document-level covariates into topic content variables, thereby affecting topic-word probability, and studying the distribution of topic words among covariates.

3 MATERIAL AND METHODS

The research framework proposed in this paper is as follows: First, use a web crawler program based on Python language to crawl user online comments and their corresponding product ratings, and complete data preprocessing; Then, use STM to extract the subject words in the user's online comments, filter the topics related to the demands, and visualize the demands; Finally, the product requirements obtained by combining subject terms with experience value and the correlation between topics are analyzed. In order to illustrate the changes and advantages brought by the introduction of covariates in STM, the Term Frequency-inverse Document Frequency (TF-IDF) algorithm was used to extract the keywords in the user's positive and negative comments, and the subject words obtained by STM. Make a comparison and analyze the impact of the frequency difference between keywords in positive reviews and negative reviews. The research framework is shown in Fig. 2.

The first step is data collection and preprocessing. A web crawler program written in Python language obtains user reviews and their ratings from experience product review sites, and performs operations such as filtering, cleaning, and de-duplication on the data. Use the ‘prepDocuments’ function to normalize and index the data, and delete comments with less than 50 words. Use the ‘TextProcessor’ function of the STM package in the R language to process the data, including the extraction of stems, the removal of default stop words and custom stop words (the, is ,at). And the deletion of numbers and punctuation to ensure that the covariate corresponding to each comment does not contain missing values for subsequent use in the STM model.

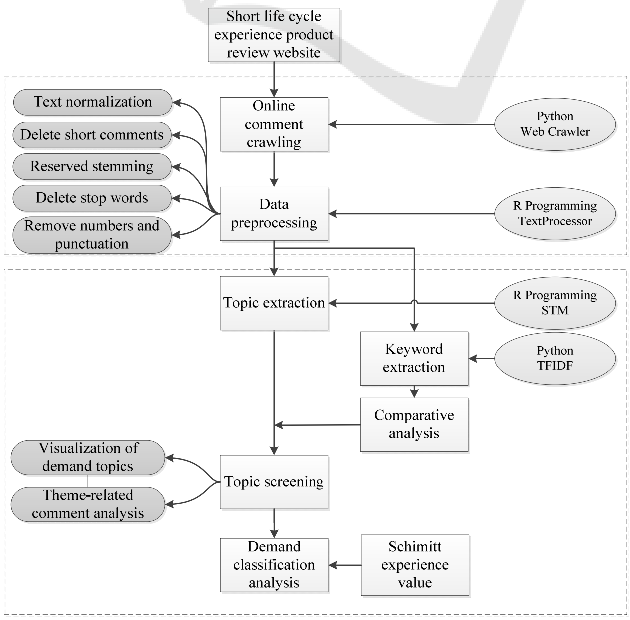

Figure 2: Research framework of demand mining method for short life cycle experience products.

The second step is demand mining and demand analysis. Use the STM model to perform text mining on the pre-processed review data, use the ratings selected by the reviewer to divide the reviews into positive reviews and negative reviews, and introduce the positive and negative of the reviews as a topic popular covariate into the STM and perform topic extraction, According to the words in the topic to manually classify and determine the topic label. Use word cloud graphs to visualize words that appear in key topics, and extract and analyze relevant comments on certain key topics. Using the TF-IDF algorithm to extract keywords for positive reviews and negative reviews, compare and analyze the extracted keywords with the subject terms extracted using the STM method, explaining the difference between STM and the conventional word frequency method for demand mining. It reflects the necessity of using STM for demand mining.

The third step is demand classification analysis. Therefore, based on the use of STM for subject extraction, experience value is added to classify user demands, to obtain user demand classification for short life experience products, and analyze each experience value classification.

4 RESULTS & DISCUSSION

This article takes the typical product movie of short life cycle experience product as an example, and conducts demand mining and demand analysis from movie user reviews.

4.1 Data Collection and Preprocessing

IMDB (Internet Movie Database, IMDB.com), as the most detailed movie database in the world, provides a platform for global movie critics to express their personal opinions on movies that have been released. Users demand to rate movies while commenting. The number represents the overall degree of user satisfaction with the movie, and reflects the user's subjective judgment on the movie. This article uses the score corresponding to the review as the standard for dividing positive reviews and negative reviews.

Using the Python language web crawler to crawl the comments from the IMDB website, the movie ratings corresponding to the comments, the reviewer ID, and the review time. This article selects the user reviews of Captain Marvel released in 2019 as the data source because the number of the reviews is reasonable, and the difference in the number of positive and negative comments is small. After deleting invalid comments, 6144 comments were finally collected, and the number of comments corresponding to each comment star rating is shown in Figure 3. Since the number of positive reviews with more than 5 stars is much higher than the number of negative reviews with less than 5 stars, in order to balance the gap in the number of positive and negative reviews, 1-6 star reviews are regarded as negative reviews and 7-10 star reviews are regarded as positive reviews. The positive and negative of is introduced as a covariate into the STM model.

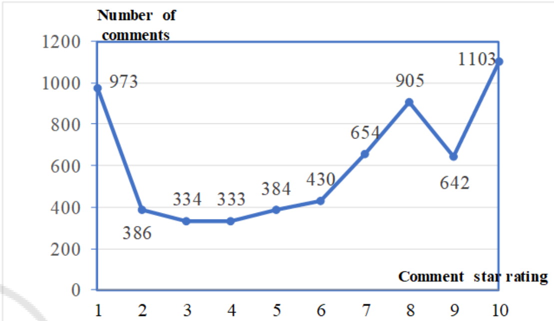

Figure 3: Star distribution of movie reviews.

4.2 Topic Extraction

Input the crawled movie reviews and their corresponding pre-processed comment positive and negative as topic popular covariates into STM, change the number of topics several times for model training, determine the optimal number of topics to be 20, and use the STM model for topic extraction. Table 1 shows the results of the extracted topics.

"Prob" represents the word with the highest occurrence probability in the topic, but the word with the highest occurrence probability in a certain topic may also appear with a high probability in another topic, so the degree of discrimination is insufficient. STM introduces the FREX (Frequency-exclusivity) statistic, which is defined as the ratio of topic-based word frequency to word-topic exclusivity, which can avoid quoting only the most commonly used words. STM also introduces the metric "Lift". This metric refers to the probability of words appearing on the topic divided by the probability of words appearing in the entire corpus. This metric will highlight the more common words in the topic than in the corpus, which the frequency of occurrence in this topic is much higher than in the entire corpus.

According to the words extracted from the above metrics, a total of 20 topic tags are designed. Among them, 20 topics related to movie features or the process of watching movies are selected: topic1

(Actor acting), Topic2 (Sexism), Topic3 (Comparative movies), topic4 (Leading actors), topic5 (Other actors), topic6 (IP animation) Topic7 (User emotion), topic8 (Feminism), topic9 (Related movies), topic10 (Movie plot), topic11 (Watching experience), topic12 (Main IP movie), topic13 (Personal thoughts), topic14 (Marvel Universe), topic15 (Personal expectations), topic16 (starring posture), topic17 (User comments), topic18 (Plot perception), topic19 (Plot trend) and topic20 (Deep thinking). See Table 1.

4.3 Positive and Negative Topics Discovery

Using the STM model to find the positive and negative tendencies of all demand topics. The results are shown in Figure 4. The demand topic with positive tendency represents the positive feedback of commenting users on the movie, while the demand topic with negative tendency represents the negative feedback of commenting users on the movie. The topics with strong tendency can be identified and extracted through STM, so as to more accurately identify the demands of consumers.

Table 1: Movie demand topic extraction results.

Topic No.	Topic Label	Criteria	Word1	Word2	Word3	Word4	Word5
1	Actor Acting	*Prob*	charact	feel	marvel	mcu	like
		FREX	phase	stake	risk	care	style
2	Sexism	*Prob*	femin	white	charact	feminist	men
		FREX	garbage	white	femin	insult	sexism
3	Comparative Movies	*Prob*	like	thing	get	make	time
		FREX	obstacle	shake	scheme	sound	mood
4	Leading Actors	*Prob*	larson	brie	charact	jackon	samuel
		FREX	samuel	jackson	perform	larson	actor
5	Other Actors	*Prob*	man	thor	black	better	war
		FREX	iron	thor	panther	man	america
6	IP Animation	*Prob*	furi	great	carol	love	nick
		FREX	carol	goos	nick	fury	danver
7	User Emotion	*Prob*	good	realli	scene	action	great
		FREX	good	funny	pretty	scene	cool
8	Feminism	*Prob*	just	female	woman	power	strong
		FREX	strong	super	women	girl	female
9	Related Movies	*Prob*	save	money	alita	battle	mcu
		FREX	save	alita	angel	battle	grace
10	Movie Plot	*Prob*	marvel	kree	earth	skrull	captain
		FREX	ver	kree	earth	ryan	boden
11	Watching Experience	*Prob*	watch	see	end	will	wait
		FREX	watch	worth	theater	cinema	must
12	Main IP Movie	*Prob*	marvel	captain	avenge	endgame	universe
		FREX	universe	avenge	captain	thano	marvel
13	Personal Thoughts	*Prob*	review	see	people	enjoy	hate
		FREX	review	negat	rate	hate	troll
14	Marvel Universe	*Prob*	marvel	captain	charact	space	point
		FREX	space	sovel	nostalg	limit	relev
15	Personal Expectations	*Prob*	high	recommend	incred	special	effect
		FREX	recommend	high	incred	balanc	blow
16	Starring Posture	*Prob*	marvel	worst	actress	ever	poor
		FREX	worst	ever	gadot	poor	shame
17	User Comments	*Prob*	bore	just	bad	charact	time
		FREX	bore	horrible	asleep	bland	uninterest
18	Plot Perception	*Prob*	polit	emake	comic	express	like
		FREX	polite	express	view	facial	shove
19	Plot Trend	*Prob*	stori	origin	mcu	marvel	great
		FREX	origin	pack	stori	nice	classic
20	Deep Thinking	*Prob*	just	felt	realli	way	think
		FREX	felt	rush	familiar	question	around

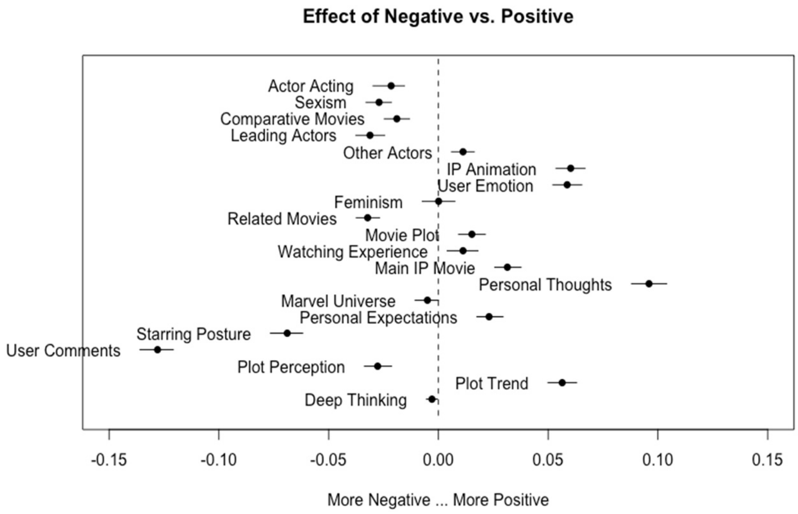

Figure 4: Movie demands thematic sentiment tendencies.

Figure 5: Cloud chart of demand subject words with positive.

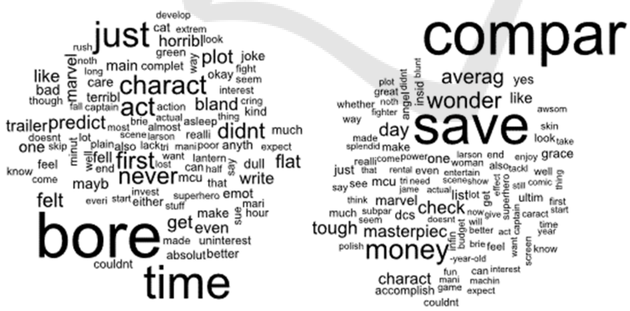

Figure 6: Cloud chart of demand subject words with negative.

A total of 4 topics with strong positive and negative tendencies are visualized by using word cloud diagram. The results are shown in Figure 5 and Figure 6. The results of the cloud chart of the subject words with positive tendency show that the comments in the personal thought topic mainly focus on the personal views of the comment users, mainly emotional words; The topic words of plot trend focus on the originality of the plot and the general story line of the movie company where the movie is located. Combined with the positive tendency of the topic, the comment users have a positive evaluation of the plot trend of the movie.

The cloud chart results of subject words with negative tendency show that the subject words evaluated by users are mainly the evaluation words used by users for the movie, mainly adjectives. The high-frequency words in the starring posture topic

Table 2: User comment keywords in positive and negative movie reviews.

Keywords (Negative)	Frequency of occurrence	Keywords (Positive)	Frequency of occurrence
marvel	0.0116	marvel	0.0122
captain	0.0053	captain	0.0062
like	0.0053	story	0.0042
character	0.0052	brie	0.0038
brie	0.0039	mcu	0.0033
larson	0.0036	watch	0.0024
story	0.0032	reviews	0.0019
mcu	0.0030	fury	0.0019
boring	0.0023	fun	0.0019
acting	0.0021	carol	0.0018
jackson	0.0015	avengers	0.0018
endgame	0.0015	origin	0.0015
samuel	0.0011	acting	0.0012

mainly describe the appearance of the heroine of the movie.

4.4 Comparative Analysis of TF-IDF Method and STM

TF-IDF algorithm is used to extract the keywords of negative comments and positive comments respectively, and the results are compared with the topics extracted by STM and their comment tendency. Since the TF-IDF value is the product of the two statistics of word frequency and inverse document frequency, it is difficult to understand its intuitive meaning. Select the effective words related to the topic from the words with high TF-IDF value to calculate their occurrence frequency in the document. The results are shown in Table 2.

Through comparative analysis, it is found that most of the negative or positive keywords extracted by TF-IDF method are consistent with the tendencies of their topics, but there are also exceptions. The keyword "endgame" extracted from the negative comments appears in the topic12 "main IP movie" with a positive tendency, and the keyword "story" appears in the topic19 "plot trend" with a positive tendency, which indicates that the reviewers' comments on the main IP movie and movie plot corresponding to the movie are mostly in the range of negative comments. However, "main IP movie" and "plot trend" are not the aspects that the commentators are dissatisfied with the movie. On the contrary, the commentators' emotional tendency towards these two topics is positive.

Accordingly, the keyword "Brie" with high TF-IDF value extracted from the positive comments, as a part of the name of the leading actor Brie Larson, appears in the topic topic4 "leading actor" which tends to be negative, indicating that the commentators also refer to the leading actor more in the positive comments with high scores, but the commentators tend to comment more negatively on the actor. This is also verified by the fact that the leading actor's surname "Larson" appears in negative comments with a high frequency, while the frequency in positive comments decreases significantly.

This shows that the frequency difference between negative comments and positive comments of key high-frequency words obtained by TF-IDF or word frequency calculation method will affect the results of demand mining. In the case of positive and negative comments with the same keyword and unable to judge the word tendency, or the original words belonging to a certain tendency are attributed to the comment keywords with the opposite tendency, lack of certain accuracy.

4.5 Demand Classification Analysis Combined with Experience Value

The topics are classified using five experience value classification methods in the strategic experience module (Bassi 2010). Classify the topic labels according to the experience value classification, and extract the positive and negative tendency of the topic corresponding to each demand topic. Table 3 shows the movie demand used in the demonstration. The movie demand is mainly composed of the experience value type, the topic label corresponding to the experience value type and the positive and negative tendency of the topic label.

Table 3: Combining movie demands with experience value types.

Type of Experience value	Value Description	Topic serial number	Topic label	Topic tendency
Sense	Affect consumers' external perception	1	Actor Acting	Negative
		4	Leading Actors	Negative
		5	Other Actors	Positive
		11	Watching Experience	Positive
		16	Starring Posture	Negative
Feel	Affect consumer sentiment	7	User Emotion	Positive
		10	Movie Plot	Positive
		17	User Comments	Negative
		18	Plot Perception	Negative
		19	Plot Trend	Positive
Think	Elicit consumers' ideas on specific topics	3	Comparative Movies	Negative
		6	IP Animation	Positive
		9	Related Movies	Negative
		12	Main IP Movie	Positive
Act	Stimulate consumers to change their lifestyles and expectations	13	Personal Thoughts	Positive
		15	Personal Expectations	Positive
		20	Deep Thinking	Negative
Realte	Arouse consumers to think about the connection with society or culture	2	Sexism	Negative
		8	Feminism	Negative
		14	Marvel Universe	Negative

5 CONCLUSIONS

This paper proposes and constructs a short life cycle experience product consumer demand mining method based on structural topic model. Using structural topic model method, the demand topics of positive and negative comments are extracted from online comments. Combined with consumer experience value, a structured experience product demand classification is constructed. It provides new ideas for improving the product quality and innovation direction of experience products. Taking a movie as an example, the effectiveness of this method is verified. Compared with the traditional keyword topic extraction method. The results show that the demand mining method proposed in this paper considers the impact of the inconsistent number of positive and negative comments in online comments, and effectively improves the accuracy of extracting demand topics from online comments, helps to provide guidance for enterprises and manufacturers developing short life cycle experience products to quickly obtain user demands, and promote the rapid innovation and generation of short life cycle experience products.

There are deficiencies in the selection of model covariates. Only the scores of online comments are considered. More online comment features or online comment user features (such as comment time, comment humanity) can be used as covariates for demand mining.

ACKNOWLEDGMENT

This study is supported by the National Natural Science Foundation of China (Grant No. 71672004).

REFERENCES

Bassi F. Experiential goods and customer satisfaction: An application to films[J]. Quality Technology & Quantitative Management, 2010, 7(1): 51-67.

Chandelier M, Steuckardt A, Mathevet R, et al. Content analysis of newspaper coverage of wolf recolonization in France using structural topic modeling[J]. Biological conservation, 2018, 220: 254-261.

Curry T A, Fix M P. May it please the twitterverse: The use of Twitter by state high court judges[J]. Journal of Information Technology & Politics, 2019, 16(4): 379-393.

Dong L, Ji S, Zhang C, et al. An unsupervised topic-sentiment joint probabilistic model for detecting

deceptive reviews[J]. Expert Systems with Applications, 2018, 114: 210-223.
Feng J, Rao Y, Xie H, et al. User group based emotion detection and topic discovery over short text[J]. World Wide Web, 2019: 1-35.
Jo Y, Oh A H. Aspect and sentiment unification model for online review analysis[C]//Proceedings of the fourth ACM international conference on Web search and data mining. 2011: 815-824.
Kuhn K D. Using structural topic modeling to identify latent topics and trends in aviation incident reports[J]. Transportation Research Part C: Emerging Technologies, 2018, 87: 105-122.
Lin C, He Y. Joint sentiment/topic model for sentiment analysis[C]//Proceedings of the 18th ACM conference on Information and knowledge management. 2009: 375-384.
Lester C A, Kessler J M, Modisett T, et al. A text mining analysis of medication quality related event reports from community pharmacies[J]. Research in Social and Administrative Pharmacy, 2019, 15(7): 845-851.
Rao Y, Li Q, Mao X, et al. Sentiment topic models for social emotion mining[J]. Information Sciences, 2014, 266: 90-100.
Roberts M E, Stewart B M, Tingley D, et al. Structural topic models for open-ended survey responses[J]. American Journal of Political Science, 2014, 58(4): 1064-1082.
Sangwan N, Malik V, Sunesh, et al. A Survey for Text Analytics Using R Language[J]. Advances in computational ences and technology, 2017, 10(7):2173-2182.
TANG Zhong-jun, CUI Jun-fu, TANG Xiao-wen, ZHU Hui-ke. A Demand Feature Pattern Mining Method for Short Life Cycle Experiential Product by Integrating Content Analysis and Association Analysis[J]. Chinese Journal of Management Science, 2019, 27(11): 166-175.
Tvinnereim E, Fløttum K. Explaining topic prevalence in answers to open-ended survey questions about climate change[J]. Nature Climate Change, 2015, 5(8): 744-747.
Usama M, Ahmad B, Song E, et al. Attention-based sentiment analysis using convolutional and recurrent neural network[J]. Future Generation Computer Systems, 2020, 113: 571-578.