



Article ID 1007-1202(2021)06-0464-09

DOI <https://doi.org/10.1051/wujns/2021266464>

Online Latent Dirichlet Allocation Model Based on Sentiment Polarity Time Series

□ HUANG Bo¹, JU Jiaji¹, CHEN Huan²,
ZHU Yimin³, LIU Jin¹, SHI Zhicai^{1,4}

1. School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2. Henan Costar Group Ltd, Nanyang 450061, Henan, China;

3. School of Computer Engineering and Science, Shanghai University, Shanghai 200244, China

4. Shanghai Key Laboratory of Integrated Administration Technologies for Information Security, Shanghai 200240, China

© Wuhan University 2021

Abstract: The Product Sensitive Online Dirichlet Allocation model (PSOLDA) proposed in this paper mainly uses the sentiment polarity of topic words in the review text to improve the accuracy of topic evolution. First, we use Latent Dirichlet Allocation (LDA) to obtain the distribution of topic words in the current time window. Second, the word2vec word vector is used as auxiliary information to determine the sentiment polarity and obtain the sentiment polarity distribution of the current topic. Finally, the sentiment polarity changes of the topics in the previous and next time window are mapped to the sentiment factors, and the distribution of topic words in the next time window is controlled through them. The experimental results show that the PSOLDA model decreases the probability distribution by 0.160 1, while Online Twitter LDA only increases by 0.069 9. The topic evolution method that integrates the sentimental information of topic words proposed in this paper is better than the traditional model.

Key words: topic evolution; sentiment factors; word vector; Latent Dirichlet Allocation (LDA)

CLC number: TP 305

Received date: 2021-07-26

Foundation item: Supported by the Opening Project of Shanghai Key Laboratory of Integrated Administration Technologies for Information Security (AGK2019004), Songjiang District Science and Technology Research Project(19SJKJGG83), National Natural Science Foundation of China(61802251)

Biography: HUANG Bo, male, Ph.D., Associate professor, research directions: software engineering, artificial intelligence. E-mail: huangbosues@sues.edu.cn

0 Introduction

The rapid development of the internet has brought important changes to people's daily life. For example, the rapid development of e-commerce has generated many commodity review texts. However, in these review texts, the topic information changes differently over time. Mining the problems and sentimental information reflected in the review text has great significance to businesses and regulatory authorities.

Topic evolution refers to the process of changes over time of the main components within the text, and tracking of the topic evolution can help researchers understand the development trend of the goal and the process of change^[1]. In recent years, academia has conducted a large number of topic evolution studies in various fields, providing a scientific reference to relevant industries and research^[2-4]. The evolution process of the theme is gradually transformed from the adjustment state to a mature state, accompanied by knowledge transfer in the evolution process, so this paper takes the emotional change as the evolution process of the external information intervention theme.

As a standard analysis method of the topic model, the Latent Dirichlet Allocation (LDA) model was proposed by Blei *et al*^[5] after the development of Latent Semantic Indexing (LSI) and Probabilistic Latent Semantic Indexing (PLSI), and then widely used in many fields such as topic mining, topic crawler, a recommendation system, and text classification. Alsumait *et al*^[6] proposed Online Latent Dirichlet Allocation (OLDA) for topic modeling of temporal text data. Lau *et al*^[7] proposed Online Twitter LDA based on OLDA, which con-

trols the distribution of topic words by introducing contribution factors to adjust the parameter's changes of the previous and next time windows. Kalyanam *et al* [8] combined the social background and text content to explain the role of adding other information on topic evolution. Hu *et al* [9] expressed the topic as a beta distribution over time and as a Dirichlet distribution about emotions and made emotion classification and confusion contrast. Chen *et al* [10] proposed the OLDA-based Forum Hot Topic Evolution Tracking Model (HTOLDA), which reduces the dimension of text from lexical space to topic space, then clusters it to find and transmit hot topics. Refs. [11-14] eliminated the time slice of documents containing old topics in the topic content matrix by constructing the topic similarity matrix. Zhang *et al* [15] obtained the vector expression of topic words according to the topic model and then obtained the evolution path of the topic through cluster analysis. As mentioned above, the existing methods mainly detect the sensitivity of topic evolution from two perspectives.

According to the rich emotional information contained in the review text, many topic sentiment hybrid models have been proposed [16,17]. Cui *et al* [18-21] proposed to construct an ideal comment set of positive and negative emotions based on the LDA model, calculate the topic similarity between real reviews and ideal reviews, and classify the sentiment of review text. Xu *et al* [22-24] obtained the expression of topic words based on topic2vec, and then calculated the content intensity and sentimental tendency of the same topic through CNN, and analyzed the evolution of the topic. An *et al* [25-28] used word2vec and k-means for topic detection and adopted a multi-source sentiment analysis method based on sentiment dictionary to analyze the co-evolution of topic and sentiment. Liu *et al* [29,30] used the sentimental information of the previous moment as the priority of the current sentimental parameters in the topic model and used cross-entropy to calculate the sentimental similarity.

As mentioned above, the topic model integrating sentiments presents the sentimental information in the text through two aspects: on the one hand, the sentimental analysis is carried out through the topic words after the topic modeling [31,32], on the other hand, sentimental information is integrated into the transmission process of prior parameters of the topic model. The current method of adding sentimental polarity in the topic evolution process is to calculate in the current time window.

Different from the method mentioned above, this paper proposes a Product Sensitive Online Dirichlet Alloca-

tion model (PSOLDA) model which uses LDA to obtain the topic word distribution in the current time window and proposes a novel topic word sentimental polarity judgment method. We introduce sentimental factors to reflect the change in sentimental polarity of the topic word in the different time windows and use sentimental factors to adjust the topic distribution of the current time window and enhance the sensitivity of the model to new topics. The experimental results show that the PSOLDA model proposed in this paper has a greater improvement in the subject detection field compared with the previous model.

1 The Framework of PSOLDA

The LDA model considers a document as a bag of words where a document is represented as a multinomial distribution over topics, and a topic is represented as a multinomial distribution over words.

The OLDA model is used to process time-series text data by inputting documents of a fixed time slice into a single LDA model. The historical probability distribution of the topic is suppressed from the prior distribution parameters of the topic as the current time slice topic. The historical probability distribution can affect the distribution of subject words in the current time window.

$$\beta_k^t = B_k^{t-1} \omega^\delta \quad (1)$$

As shown in formula (1), the prior parameters of the subject word are preserved in the parameter evolution matrix. B_k^{t-1} represents the evolution matrix of the k topics on the δ time slices. δ is the length of the historical time slice. ω^δ is the weight on the different time slice, and the different size of the weights determines the different impact of the historical time slice on the current slice. β_k^t is the prior parameter of the topic k in a time slice t . The OLDA model determines the evolution of the topic by setting confidence values and comparing the parameter values of the forward and backward matrices.

The Online Twitter LDA model is mainly used to discover hot topics in short texts, such as Twitter and Weibo. The advantage of this model is that the model's parameters will not grow with the new input text. Online Twitter LDA divides the text into several time slices according to the time series and then slides a fixed-size time window to keep the number of time slices in the time window unchanged. Compared with OLDA, it has a fixed-size parameter matrix, so it has a higher sensitivity to topic detection. Lau *et al* [7] proposed to use contribut-

ing factors to determine the relationship between the parameters of the previous and the next time windows. The parameters of the previous and next time windows are defined as follows.

Text for the previous time window:

$$\alpha'_{dt} = \frac{n(d,t)}{N_{old}} \times D_{old} \times T \times \alpha \quad (2)$$

$$\beta'_{tw} = \beta_0 \times (1-c) + \frac{n(t,w)}{N_{old}} \times D_{old} \times T \times W_{new} \times \beta_0 \times c \quad (3)$$

Text for the new arrival time window:

$$\alpha'_{dt} = \alpha_0; \beta'_{tw} = \beta_0 \quad (4)$$

Among them α'_{dt} is the priority of Topic t in Text d ; β'_{tw} is the priority of word w in Topic t ; $n(d,t)$ is the number of Topics t in the text d ; $n(t,w)$ is the number of Word w in Topic t . α_0 and β_0 are the initial prior values, D_{old} is the number of texts; T is the number of Topic, N_{old} is the number of Word in the previous time window, and W_{new} is the size of a new dictionary, and we can normalize formula (2),(3),(4) by formula (5), (6).

$$\sum \alpha' = \sum \alpha = D \times T \times \alpha_0 \quad (5)$$

$$\sum \beta' = \sum \beta = T \times W \times \beta_0 \quad (6)$$

To study the topic evolution of commodity review text, this paper proposes the PSOLDA model based on Online Twitter LDA, which integrates the sentimental features of the topic. Different from Online Twitter LDA, this model introduces the sentimental polarity of the topic and the sentiment changes into the topic evolution of the previous and next time windows. A contributing factor c is defined to reflect changes in the evolution of the topic. The model is mainly composed of three parts: the first part is text preprocessing, which mainly includes word segmentation and stop words, as well as word2vec word vector training; The second part is to model the text of the current time window with LDA and calculate the sentimental polarity of the topic; The third part is to integrate the results of topic sentiment calculation into the process of the topic evolution. The model framework is shown in Fig.1.

2 Topic Evolution under the Control of Sentiment Polarity

2.1 Improved Topic Sentimental Polarity Algorithm

There are two scenarios when calculating the sentimental polarity of topic words: when the topic words

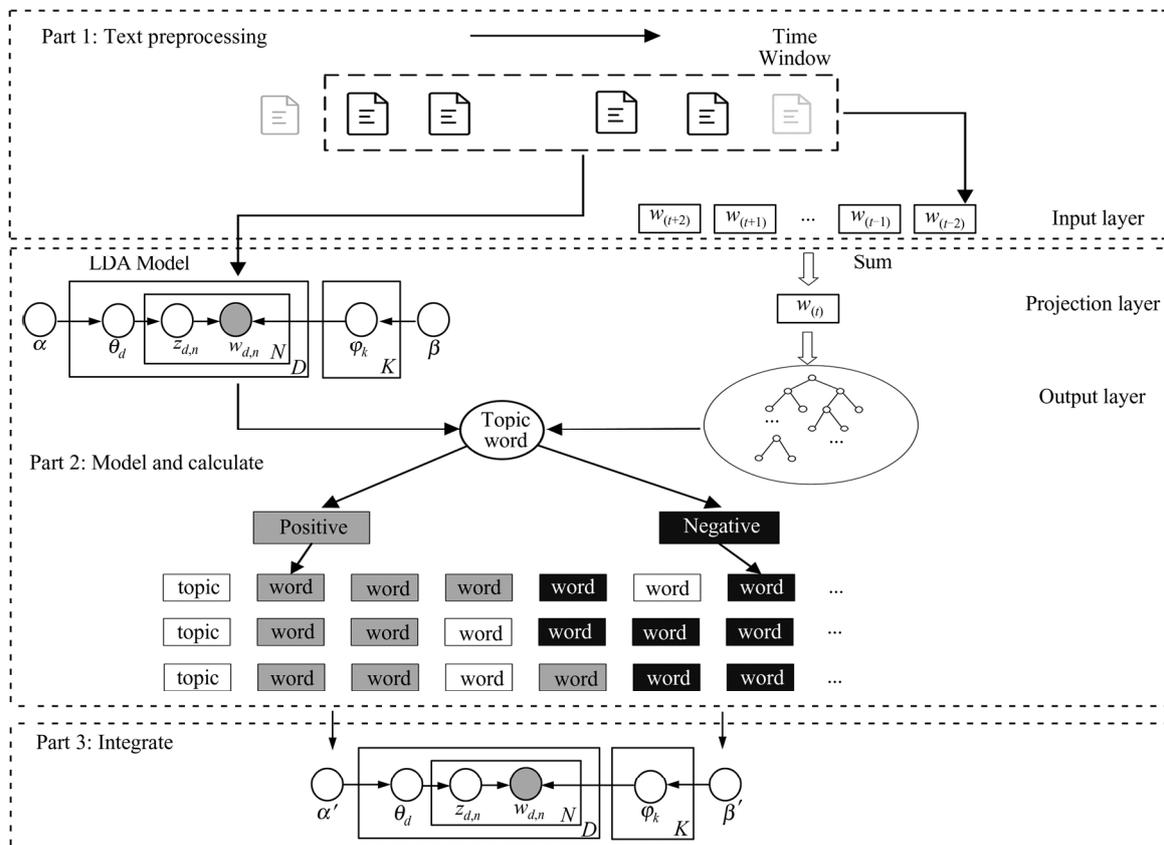


Fig.1 PSOLDA frame diagram

exist in the sentiment dictionary, query directly the sentiment dictionary to calculate the sentimental polarity; When the topic word does not exist in the sentiment dictionary, the similarity of the topic word can be used for calculation. However, there are situations that there is a high degree of similarity between the two words, but the sentimental polarities of these words are opposite. For example, a certain topic word is “good”, and the words closest to the topic word in the word vector space can be obtained by using the word vector as [“satisfactory”, “poor”, “bullish”, “awesome”, “regular”, “surprised”, “nice”, “legendary”, “humanized” and “stable”]. We aim to construct a topic sentiment polarity classification algorithm to avoid having a high similarity of a word but opposite sentimental polarity. When calculating the sentimental polarity of topic words, similar phrases are constructed for topic words that are not in the sentiment dictionary, and the sentimental polarity of words in similar phrases is judged to obtain the sentimental polarity of topic words.

Define $\varphi_z = (w_1, w_2, w_3, \dots, w_n)$ as the word distribution of Topic z , n as the number of topic words, $[w'_{i1}, w'_{i2}, w'_{i3}, \dots, w'_{in}]$ as a similar phrase obtained by using word vector similarity for the i -th word in the topic word distribution, in which n_w is the number of similar words required. The formula for calculating Topic z sentimental polarity value is:

$$S = \sum_{i=1}^n \left(w_h w_i x_i + \left(\sum_{j=1}^n w_i w'_{ij} x'_i \right) / n \right) \quad (7)$$

If the topic word w_i is positive or negative in a sentiment dictionary, then $x_i = \pm 1$ and $x'_i = 0$; If the topic word w_i is not in a positive or negative sentiment dictionary, then $x_i = 0$ and $x'_i = \pm 1$; w_h is used to adjust the weight of topic words in the sentiment dictionary when we determine the sentimental polarity of the topic. If $S \geq 0$, then the topic is determined to be a positive sentimental topic; If $S < 0$, then the topic is determined to be a negative sentimental topic.

2.2 The Calculation Method of Sentiment Factors

To study the effect of topic sentiment change on topic evolution, a topic evolution model integrating topic sentiment polarity is proposed based on the research of Lau *et al*^[7]. Lau proposed to control the influence of the parameters of the previous time window on the current time window information by changing the contribution factor. Different from the method proposed by Lau *et al*^[7], this paper maps changes in sentimental polarity of the topic to changes of sentiment factors, then obtains the prior parameters α, β of the previous time window to

the current time window through sentimental factors.

Define the number of positive sentiment topics in the current time window as n_{pos} , the number of negative sentiment topics as n_{neg} , the number of positive topics in the previous time window as n'_{pos} , the number of negative sentiment topics as n'_{neg} , K as the number of topics modeled for LDA. The formula for calculating sentiment factor c is as follows:

$$c = 1 - \text{sigmoid}' \left(\frac{|n_{pos} - n'_{pos}| + |n_{neg} - n'_{neg}|}{K} \right) \quad (8)$$

$$\text{sigmoid}' = \frac{1}{1 + e^{-12x}}, \quad x \in [0, 1] \quad (9)$$

The curve of improved sigmoid' function is shown in Fig.2, the function domain is $[0, 1]$, and the range of function values is $[0.5, 1]$. Calculate the number of changes of positive sentimental topics and negative sentimental topics in the previous and next time windows and normalize them. Then use the sigmoid' function to map the change of topic sentiment to the change within the value range of sigmoid' function.

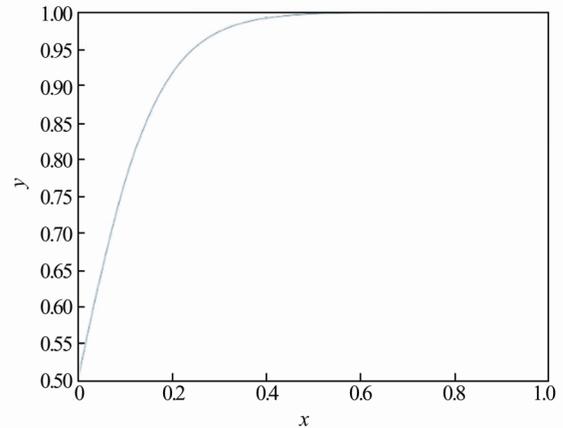


Fig.2 Improved sigmoid' function curve

2.3 The Topic Evolution Algorithm

The topic evolution algorithm integrating sentiment factors is expressed as follows:

First, read in the l time slice documents of the first time window, input it to the word2vec model for word vector training and input it to the LDA model to obtain the topic word distribution. Second, calculate the topic sentimental polarity distribution S_0 for the first time window. Then update the document and the word2vec model through sliding the time window, and calculate the topic sentimental factor according to the topic sentimental polarity distribution of the previous and the next time windows. Finally, calculate the topic

distribution θ, φ at the current time according to the topic sentimental factor α', β' through Gibbs sampling keyword distribution of the window.

Algorithm PSOLDA Specific Process

Input: α_0, β_0, K, l

Output: θ, φ

1. Read in the document sets in l time slices and segment words
2. Get $G_0 = [k_0, k_1, \dots, k_{l-1}]$
3. Update word2vec model // Training word vector model
4. Run LDA for $[k_0, k_1, \dots, k_{l-1}]$
5. Compute S_0 //Calculate the sentimental polarity distribution in the first time window
6. **For** k_1 to k_l , remove the oldest time slice document and add a new time slice document
7. Update G
8. Update word2vec model
9. Run LDA according G
10. Compute according to formula (4)
11. Compute c according to formula (5)
12. Calculate α', β' according to c
13. Gibbs sampling according to α', β' //Obtain a stable probability distribution.
14. Run Gibbs(α', β')
15. Get θ, φ

End For

3 Experiment and Result Analysis

3.1 Dataset Acquisition and Sentiment Dictionary Construction

The experimental data was crawled from the online reviews on financial products of the Online Lending House. A total of 28 772 reviews on seven financial products from September 2017 to March 2019 were collected through writing crawler scripts. The obtained review information is sorted by time series, and time slices are divided by month.

Among them, the number of comments in 2018 varies by month as expressed in Fig.3. It can be seen that under normal circumstances, the number of comments has been stable below 2 000, but in July and August, the number of comments increased significantly. It might be related to the deterioration of the overall P2P investment environment during this period, such as the explosion of some platforms and the strengthening of the country's control of P2P platforms, which caused investors to worry about their investment.

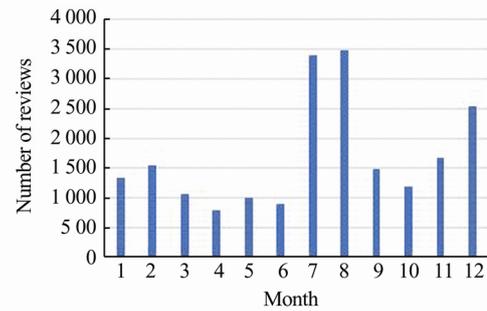


Fig.3 Distribution of reviews in 2018

LDA algorithm is used to obtain the distribution of topic words for all texts, and the first 15 words of each topic word distribution are used to represent the topics. Then we get the 10 words that are closest to each topic word through word2vec, for a total of 6 000 words. Through the manual annotation, 284 words in the positive sentiment dictionary and 338 words in the negative sentiment dictionary are obtained, respectively.

3.2 Determination of Optimal Parameters

To obtain the optimal number of topics, LDA is used to model all review data sets and calculate the perplexity. Perplexity is mainly used to measure the prediction ability of unknown data. The smaller the perplexity, the stronger the prediction ability of the model. Perplexity is defined as follows:

$$P(\text{test}) = \exp \frac{\sum_{d=1}^M \sum_{o=1}^d \ln p(w_d^j)}{\sum_{d=1}^M N_d} \quad (10)$$

The denominator is the sum of all words in the test set, namely the total length of the test set, and $p(w_d^j)$ is the probability of each word in the test set.

Set $\alpha_0 = 0.625$, $\beta_0 = 0.01$, input the daily review texts as a time slice into the LDA model to obtain the K fitting curve of the perplexity with the number of topics. As described in Fig.4, when the number of topics is 40, the perplexity is the lowest, so we set the number of topics to 40.

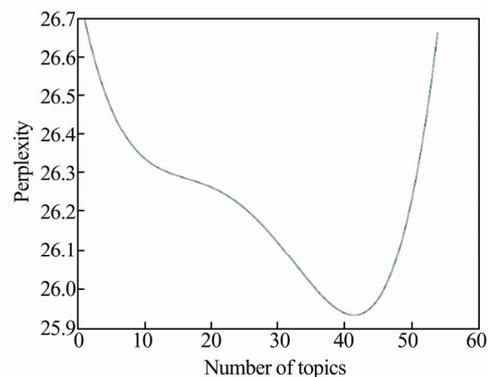


Fig.4 The curve of topics and perplexity

The accuracy of the topic classification algorithm affects the reliability of the model. The weight w_h of words in the sentiment dictionary and the number of similar words n_w that are not in the sentiment dictionary both affect the accuracy of topic sentimental polarity classification. Fixed n_w values are 1, 2, 3, 4, the accuracy of the sentiment polarity classification is shown in Fig.5.

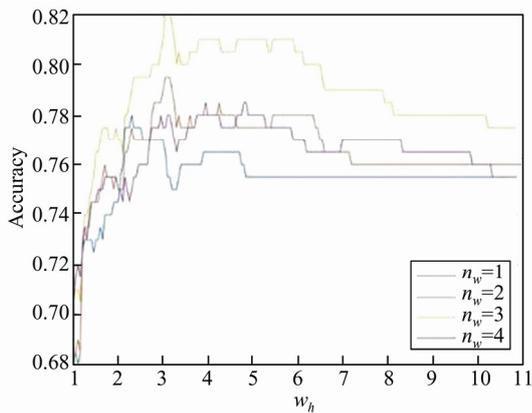


Fig.5 The curve of the weights of word and algorithm accuracy

It can be seen from Fig.5, when $n_w = 3$ and $w_h = 3.15$, the sentiment polarity classification algorithm has the highest accuracy.

Comparison of accuracy of different topic sentiment polarity classification algorithms is shown in Table 1.

Table 1 Comparison of accuracy of different topic sentiment polarity classification algorithm

Method	Accuracy
Sentiment Dictionary	0.685
Similarity	0.760
The method in this paper	0.820

3.3 Accuracy Analysis of Topic Sentiment Recognition

In order to verify the impact of topic sentiment changes on topic evolution, the time slice text in units of months is input into the LDA model, and the topic sentimental polarity distribution in each time window is obtained, as illustrated in Fig.6. The time window is set to 10, namely a time window contains text data of 10 days. In each time window of Fig.6, the yellow part in the upper half means that the sentiment polarity is negative, and the blue part in the lower half is positive. Obviously, with the change of time, the sentimental tendency of the topic words has also changed. The above information is

integrated into the model proposed in the paper to improve the accuracy of topic detection.

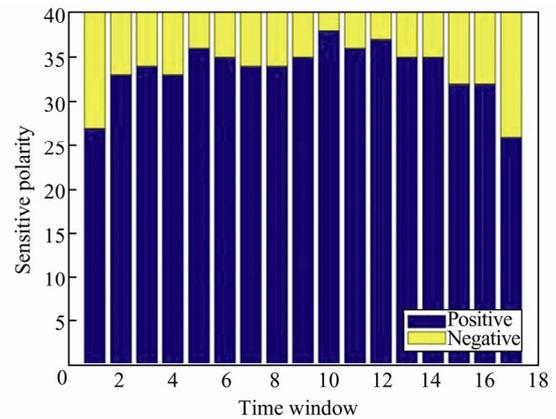


Fig.6 The distribution of topic sentiment polarity in different time windows

3.4 Comparison of Perplexity of Topic Models

The sentimental factors are calculated based on the number of topic sentiment changes in the previous and next time windows. The sentiment factor is used as the input of the topic evolution model to obtain the topic word distribution in different time windows. In this paper, perplexity of formula (10) is taken as the evaluation index of the model. The comparison of the perplexity of Online Twitter LDA in the 17-time windows is shown in Fig.7. From Fig.7, the perplexity of the PSOLDA model is smaller than that of the Online Twitter LDA. The smaller the perplexity, the stronger the prediction ability of the model. The PSOLDA significantly outperforms the original Online Twitter LDA in the time window 2-13. This results illustrate the effectiveness of the PSOLDA model.

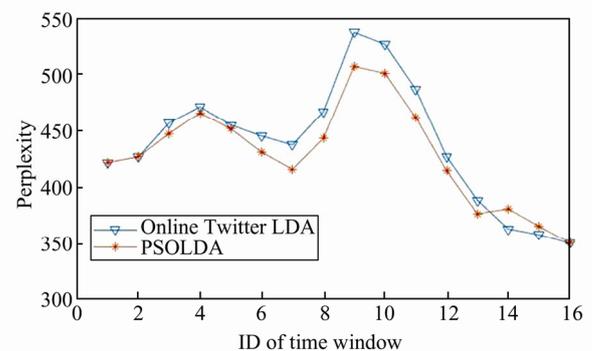


Fig.7 Comparison of perplexity in different time windows

3.5 Topic Words Evolution Analysis

The investment risk of financial product investment has always been the most concerned issue of investors. Therefore, investment risk was chosen to analyze the

evolution of topic in this paper. The text data divided according to time slices are input into the PSOLDA model to obtain the topic word distribution of the dataset in the time series, and then the investment risk is selected for topic evolution analysis in the two models. The online Twitter LDA model is selected as the baseline model for comparative analysis of the topic evolution.

It is observed from Table 2 and Fig.6 that the distribution of the topic words changed with the polarity of the topic sentiment. Among them, the topic sentimental polarity of time windows 1-2 changes considerably. Therefore, this paper elaborates on the data in Table 2 for the time window 1-2. The positive sentiment of the topic words has increased in the time widows 1-2, so the probability dis-

tribution of investment risk should be reduced.

The probability distributions of investment risk topic words obtained from the PSOLDA model were 0.524 8 and 0.364 7 in time windows 1-2. And the probability distribution of investment risk topic words from the Online Twitter LDA were 0.395 2 and 0.465 1, respectively. The PSOLDA model reduced the probability distribution of investment risk by 0.160 1, while the Online Twitter LDA increased by 0.069 9. Clearly, the results obtained by PSOLDA model are more fitting with expectation.

In Fig.6, the positive sentiment of the topic words raised in the time windows 1-5, indicating that people's attitude towards investment risk were mostly positive.

Table 2 Comparison of the distribution probability of topic words

Time window number	PSOLDA		Online Twitter LDA	
	Topic words	Probability	Topic words	Probability
Time window 1	Investment Risk	0.524 8	Investment Risk	0.395 2
	Workdays	0.035 8	Phoenix TV	0.042 4
	Registration No.	0.029 9	Bank Card	0.021 2
	Departure Point	0.017 9	Huaxia Bank	0.021 2
	Security Coefficient	0.017 9	To be Improved	0.021 2
	Full Transparency	0.012 0	Explanation	0.014 2
Time window 2	Investment Risk	0.364 7	Investment Risk	0.465 1
	Financial Products	0.113 8	Phoenix TV	0.053 2
	Listed Companies	0.056 9	To be improved	0.018 3
	Alipay	0.033 2	Steady Growth	0.017 5
	Steady Development	0.028 4	Bank Card	0.012 7
	Cautious	0.009 5	Very Important	0.012 2
Time window 3	Investment Risk	0.313 2	Investment Risk	0.485 5
	Financial Products	0.190 2	Phoenix TV	0.042 2
	P2P	0.059 9	Steady Growth	0.010 7
	Listed Companies	0.043 1	Slower	0.010 0
	Alipay 1	0.039 0	Not too bad	0.010 0
	Steady Development	0.023 5	Very Important	0.007 1
Time window 4	Investment Risk	0.233 9	Investment Risk	0.429 6
	Financial Products	0.131 3	Phoenix TV	0.046 5
	Pretty Good	0.059 2	What Happened	0.014 6
	Listed Companies	0.050 7	Steady Growth	0.011 6
	Alipay	0.050 5	Not too bad	0.011 6
	P2P	0.041 0	Too Expensive	0.011 4
Time window 5	Investment Risk	0.188 8	Investment Risk	0.399 7
	Financial Products	0.108 8	Phoenix TV	0.046 7
	P2P	0.088 7	What Happened	0.027 2
	Pretty Good	0.087 1	Steady Growth	0.013 9
	Listed Companies	0.037 1	Too Expensive	0.013 9
	Alipay	0.035 1	Very Important	0.009 3

Therefore, the probability distribution of investment risk should be gradually reduced in the expectation. In Fig.8, the probability distribution of investment risk by PSOLDA declined, while in Online Twitter LDA, its probability firstly increased and then decreased. Compared Fig.6 and Fig.8, it can be noted that the probability distribution of investment risk by PSOLDA is negatively correlated with the distribution of positive sentiment polarity for the time window 1-5. The experimental results prove that the PSOLDA model can influence the evolution of topic words from the perspective of topic sentimental polarity compared with traditional methods.

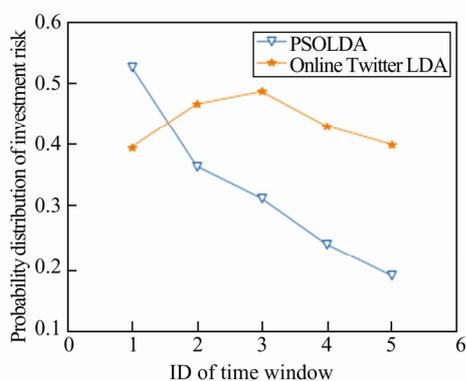


Fig.8 The probability distribution of investment risk in different time windows

4 Conclusion

The current topic evolution model that integrates emotions has the problem that the parameter matrix increases with time while the detection sensitivity decreases. Aiming at the problem, this paper proposes an improved Online Twitter LDA model based on Online Twitter LDA. Firstly, introduce the word2vec model to calculate the dynamic change of topic sentimental polarity. Then integrate the change of topic emotion polarity into the process of topic evolution. Finally, realize the dynamic evolution analysis of topic word distribution with a topic sentiment. Experiments demonstrate that the classification accuracy of the improved topic sentiment polarity algorithm is higher than that of the sentiment polarity classification algorithm based on the sentiment dictionary or similarity. The improved topic evolution model is better than the original model in terms of perplexity. And the dynamic change of sentiment factors dynamically affects the topic word distribution of the topic evolution model.

However, the method proposed in this paper also

has certain limitations. The review text is input into the time window in the form of time slices, which makes it impossible to judge the topic distribution of each review text. This will make it hard to examine the effect of the model from the perspective of text classification. Therefore, further research will focus on how to integrate the improved model into the process of short text classification.

References

- [1] Amoualian H, Clausel M, Gaussier E, *et al.* Streaming-LDA: A copula-based approach to modeling topic dependencies in document streams [C]// *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM, 2016: 695-704.
- [2] Lin Y. Topic evolution of innovation academic researches [J]. *Journal of Small Business Strategy*, 2016, **26**(1): 25-36.
- [3] Wu Y, Jin X, Xue Y. Evaluation of research topic evolution in psychiatry using co-word analysis [J]. *Medicine*, 2017, **96**(25): e7349.
- [4] Wu Q, Zhang C, Hong Q, *et al.* Topic evolution based on LDA and HMM and its application in stem cell research [J]. *Journal of Information Science*, 2014, **40**(5): 611-620.
- [5] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation [J]. *The Journal of Machine Learning Research*, 2003, **3**: 993-1022.
- [6] Alsumait L, Barbará D, Domeniconi C. On-Line LDA: Adaptive topic models for mining text streams with applications to topic detection and tracking [C]// *Proceedings of the 8th IEEE International Conference on Data Mining*. Washington D C: IEEE, 2008: 3-12.
- [7] Lau J H, Collier N, Baldwin T. On-line trend analysis with topic models: #Twitter trends detection topic model online [C]// *Proceedings of COLING 2012*. 2012: 1519-1534.
- [8] Kalyanam J, Mantrach A, Saeztrumper D, *et al.* Leveraging social context for modeling topic evolution [C]// *Knowledge Discovery and Data Mining*. New York: ACM, 2015: 517-526.
- [9] Hu Y, Xu X F, Li L, *et al.* Analyzing topic-sentiment and topic evolution over time from social media [C]// *International Conference on Knowledge Science, Engineering and Management*. Berlin: Springer-Verlag, 2016: 97-109.
- [10] Chem X S, Gao Y, Jiang H, *et al.* OLDA-based model for hot topic evolution and tracking [J]. *Journal of South China University of Technology (Natural Science Edition)*, 2016, **44**: 130-136.
- [11] Pei K F, Chen Y Z, Ma J. Variable online theme evolution

- model based on OLDA [J]. *Information Science*, 2017, **35**: 63-68.
- [12] Nimala K, Jebakumar R. Sentiment topic emotion model on students feedback for educational benefits and practices [J]. *Behaviour & Information Technology*, 2019, **38**: 1259-1272.
- [13] Xu Y, Li Y, Liang Y, *et al.* Topic-sentiment evolution over time: A manifold learning-based model for online news [J]. *Journal of Intelligent Information Systems*, 2020, **55**(1): 27-49.
- [14] Pergola G, Gui L, He Y L. TDAM: A topic-dependent attention model for sentiment analysis [J]. *Information Processing & Management*, 2019, **56**: 1325-1335.
- [15] Zhang P Y, Liu D S. Topic evolutionary analysis of short text based on word vector and BTM [J]. *Data Analysis and Knowledge Discovery*, 2019, **3**: 95-101.
- [16] Zhao R Y, Zhang Y. Evolution study of sentiment analysis based on bibliometrics of time and space dimensions [J]. *Information Science*, 2018, **36**: 171-177.
- [17] Rao Y. Contextual sentiment topic model for adaptive social emotion classification [J]. *IEEE Intelligent Systems*, 2015, **31**(1): 41-47.
- [18] Cui X L, Narisa, Liu X J. Sentiment analysis of online reviews based on topic similarity [J]. *Journal of Systems & Management*, 2018, **27**: 821-827.
- [19] Jeyaraj A, Zadeh A H. Evolution of information systems research: Insights from topic modeling [J]. *Information & Management*, 2020, **57**: 6-17.
- [20] Wu C, Kanoulas E, de Rijke M. Learning entity-centric document representations using an entity facet topic model [J]. *Information Processing & Management*, 2020, **57**(3): 102216-102278.
- [21] Zhang Y, Wei H, Ran Y, *et al.* Drawing openness to experience from user generated contents: An interpretable data-driven topic modeling approach [J]. *Expert Systems with Applications*, 2020, **144**: 113073-113080.
- [22] Xu Y M, Lü S N, Cai L Q, *et al.* Analyzing news topic evolution with convolutional neural networks and Topic2vec [J]. *Data Analysis and Knowledge Discovery*, 2018, **2**: 31-41.
- [23] Jelodar H, Wang Y L, Yuan C, *et al.* Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey [J]. *Multimedia Tools and Applications*, 2019, **78**: 15169-15211.
- [24] Wang G, Chi Y, Liu Y, *et al.* Studies on a multidimensional public opinion network model and its topic detection algorithm [J]. *Information Processing & Management*, 2019, **56**(3): 584-608.
- [25] An L, Wu L. An integrated analysis of topical and emotional evolution of microblog public opinions on public emergencies [J]. *Library and Information Service*, 2017, **61**: 120-129.
- [26] Lin P, Jiang S, Li D, *et al.* Comprehending international important Ramsar wetland documents using latent semantic topic model in kernel space [J]. *Natural Resource Modeling*, 2019, **32**(4): e12215.
- [27] Ali F, Kwak D, Khan P, *et al.* Transportation sentiment analysis using word embedding and ontology-based topic modeling [J]. *Knowledge-Based Systems*, 2019, **174**: 27-42.
- [28] Liang Q, Wang K B. Monitoring of user-generated reviews via a sequential reverse joint sentiment-topic model [J]. *Quality and Reliability Engineering International*, 2019, **35**: 1180-1199.
- [29] Liu Y W, Liu Y H, Yang S, *et al.* OTSCM approach for tracking on-line sentiment of topic [J]. *Journal of Modern Information*, 2017, **37**: 35-41.
- [30] Wang Y, Taylor J E. DUET: Data-driven approach based on latent Dirichlet allocation topic modeling [J]. *Journal of Computing in Civil Engineering*, 2019, **33**: 425-437.
- [31] Nimala K, Jebakumar R. A robust user sentiment biterm topic mixture model based on user aggregation strategy to avoid data sparsity for short text [J]. *Journal of Medical Systems*, 2019, **43**: 93-98.
- [32] Deng D, Jing L P, Yu J, *et al.* Sentiment lexicon construction with hierarchical supervision topic model [C]// *IEEE/ACM Transactions on Audio Speech and Language Processing*. New York: ACM, 2019, **27**: 704-718.

□