

## Article

# Topic-Specific Emotion Mining Model for Online Comments

Xiangfeng Luo <sup>1</sup> and Yawen Yi <sup>2,\*</sup>

<sup>1</sup> Shanghai Institute for Advanced Communication and Data Science, School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China; luoxf@shu.edu.cn

<sup>2</sup> School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China

\* Correspondence: yywen@i.shu.edu.cn

Received: 31 January 2019; Accepted: 21 March 2019; Published: 24 March 2019



**Abstract:** Nowadays, massive texts are generated on the web, which contain a variety of viewpoints, attitudes, and emotions for products and services. Subjective information mining of online comments is vital for enterprises to improve their products or services and for consumers to make purchase decisions. Various effective methods, the mainstream one of which is the topic model, have been put forward to solve this problem. Although most of topic models can mine the topic-level emotion of the product comments, they do not consider interword relations and the number of topics determined adaptively, which leads to poor comprehensibility, high time requirement, and low accuracy. To solve the above problems, this paper proposes an unsupervised Topic-Specific Emotion Mining Model (TSEM), which adds corresponding relationship between aspect words and opinion words to express comments as a bag of aspect–opinion pairs. On one hand, the rich semantic information obtained by adding interword relationship can enhance the comprehensibility of results. On the other hand, text dimensions reduced by adding relationships can cut the computation time. In addition, the number of topics in our model is adaptively determined by calculating perplexity to improve the emotion accuracy of the topic level. Our experiments using Taobao commodity comments achieve better results than baseline models in terms of accuracy, computation time, and comprehensibility. Therefore, our proposed model can be effectively applied to online comment emotion mining tasks.

**Keywords:** topic-specific emotion mining; online comment; topic model

## 1. Introduction

As time goes on, consumers are increasingly dependent on online shopping [1]. Simultaneously, the scale of transaction and size of users in the online shopping market have continued to expand. Accordingly, there are plenty of shopping websites, such as Taobao, Jingdong, Amazon, and so on. Almost all e-commerce platforms provide users with the function of online comments to express their views and feelings after receiving their goods. So far, each online shopping platform has accumulated enormous numbers of comments [2]. Online comments that mingle opinions, emotions, aspects, and time stamps contain rich and valuable knowledge [3]. An example of online comment is shown as Figure 1. In general, what's useful to us are the comment content and the timestamp.

Emotion mining for online comments certainly has been an attractive topic of study [4,5] and has a wide range of applications, such as product improvement [6] and product recommendation [7]. Through the excavation of commentary emotion, sellers can grasp buyers' attitude on products and services to make better marketing strategies. Consumers can get a reference to pick the most favored products.

Currently, the methods for sentiment analysis are based on lexicon [8] and machine learning [9]. These methods mainly concentrate on document-level emotion mining. When we need to understand

emotion of a specific topic for a product, such as the screen of the phone, it is not enough. To discover a particular topic's emotion of the entity, many researchers extend Latent Dirichlet Allocation (LDA) [10] by incorporating a sentiment layer [11,12]. Topic models based on LDA are capable of mining effectively the topic-level emotions to a certain extent. Nevertheless, these bag-of-words models do not adaptively determine the number of topics and ignore the correspondence between opinions and aspects. Uncertainty in the number of topics inevitably reduces the accuracy of topic-level sentiment analysis. A single word contains less semantic information and its emotion is ambiguous. Thus, low accuracy of emotion mining and poor interpretability of results can be caused by ignoring interword relationships. Meanwhile, the dimension of the comments that is represented by a bag of words is higher, which leads to time-consuming calculations.



Figure 1. Online commodity evaluation.

To overcome the boundedness of the conventional topic models, this paper puts forward a topic model named Topic-specific Emotion Mining Model (TSEM), which not only incorporates relationships between words, but also automatically gets the number of topics through the perplexity [13] indicator. The generation model can be described as follows. Firstly, an aspect–opinion pair's latent topic is chosen from time-specific topic distribution. The time of each review text is known. Then the aspect–opinion pair's sentiment is generated from time-topic emotional distribution under the condition of topic label and time label of the word pair ascertained. Secondly, an aspect–opinion pair is sampled from emotion-specific and topic-specific word pair distribution after getting the emotion and topic of the pair.

Because users' sentiment is uncomplicated in commodity reviews, it is not necessary to divide emotion into multiple dimensions. Therefore, we only divide the sentiment of online comments into positive and negative emotions.

The rest of this paper is arranged as below. We show the related work in Section 2. Section 3 introduces our topic model for online comments in details. Section 4 is the experimental setting and the analysis of the experimental results. At last, we make a conclusion in Section 5.

## 2. Related Work

In this section, the related work of emotion mining for online comments is firstly reviewed. Then, we will introduce the existing work of the topic-level emotion mining in this field.

Traditional online review emotion mining methods focus on machine learning [4,9,14], lexicon [8,15,16], and mixed methods [17,18].

Ye [9] applied Naïve Bayes, Support Vector Machine (SVM), and the character-based N-gram model to online review emotion mining, and experiments showed that SVM and the N-gram model have superior classification effects. Mulkalwar [14] used the classifier combine rule to combine emotional classification results of classifier Support Vector Machine and Hidden Markov Model. Tripathy [4] employed four different machine learning algorithms to classify emotions, including Naive Bayes, Maximum Entropy, Stochastic Gradient Descent, and Support Vector Machine. Dang [8] developed a lexicon-enhanced method, which used sentiment words based on a sentiment lexicon to set up a new feature dimension. Augustyniak [15] compared the two methods of emotion mining: lexicon-based and supervised learning, and proved that the lexicon-based method had a better performance in emotion

classification. Jararweh [16] presented different lexicon construction techniques and assigned the most frequent polarity label to the corresponding aspect category. Kennedy [17] proposed two methods: emotion classification based on the number of emotional words and using Support Vector Machines. Then, combining the two methods improved the performance of results. Prabowo [18] combined rule-based classification and machine learning to construct different emotion classifiers.

However, when using machine learning for sentiment classification, most of the existing methods are mainly based on supervised learning models, which are trained from tagged corpora. That is to say, each document in the corpora should be marked with emotional classification (e.g., positive or negative) before being trained. Such tagged corpora are commonly difficult to obtain in practical application. When applying lexicon-based methods to sentiment classification, most existing methods classify the emotions of a text through the emotions of words in the lexicon, which neglects the context of words. Once the emotions of words change with the context, it will inevitably lead to a large deviation between the result of emotional classification and the actual emotion of the text. In addition, when classifying text into the more fine-grained emotions (e.g., emotional classification for the topic “price” of a product), the above-mentioned approaches are often necessary to first detect the topic/feature and then categorize the emotion for that specific topic/feature.

Besides the above-mentioned methods, there are other methods for topic-level emotion mining. Topic-level emotion mining contains two subtasks: aspect-based topic extraction [19–21] and topic-based emotion classification [22]. Further attempts are mainly concentrated on completing one of the two subtasks.

Zhu [19] used a multi-aspect bootstrapping method for aspect identification. Poria [20] proposed a rule-based approach to detect aspects by exploiting common-sense knowledge and sentence dependency trees. Schouten [21] proposed an algorithm to compute a score for each implicit aspect and introduced a threshold parameter to filter out aspects whose score is too low. Jeyapriya [22] used frequent item set mining to extract aspect and utilized supervised learning algorithms to identify emotion of each aspect. Bhoir [23] used a rule-based approach to find feature–opinion pairs and determined emotion of opinion based on lexicon.

Sentiment classification is closely related to topics or domains. Intuitively, it is feasible to combine the above two subtasks to classify emotions at the topic level, but it is a little troublesome. No matter which subtask is not completed well, it will bring poor performance of the final emotional classification results. So, identifying both topic and sentiment simultaneously should play a vital role in topic-level emotion mining.

With all this in mind, the probabilistic generative model will be a good choice in mining topic-level emotion, which can achieve the goal of detecting topic and emotion at the same time. Therefore, many scholars focus their attention on the topic model to mine the topic-level emotions. Extended models based on topic model are also widely applied to online comment emotion mining, which can effectively analyze online comments’ emotions at topic level.

Mei [24] built the Topic-Sentiment Mixture (TSM) model, which separated sentiment model from topic model resulting in a lack of connection between emotions and topics. Titov and Bagheri [25] introduced a Markov chain to extend LDA and named the model ADM-LDA that was able to extract multiword aspects. However, this model does not take into account the emotion of words. Lin [11] proposed Joint Sentiment/Topic (JST) model to identify topics and emotions from texts simultaneously. Yohan [12] firstly proposed Sentence-LDA (SLDA), which assumed that all words in a sentence only belonged to one topic, and then extended SLDA to Aspect and Sentiment Unification Model (ASUM). Joint Aspect-based Sentiment Topic (JAST) model [26] can solve multiple subproblems of aspect-level sentiment analysis by modeling these subproblems under a unified framework. But it only contains limited semantic information. Park [27] proposed the Specification Latent Dirichlet Allocation (SpecLDA) model to mine comment texts relevant to a feature value, which combined product reviews and specifications that were composed of structured information. However, the acquisition of specifications is not easy.

JST and ASUM are the most similar methods to ours. ASUM assumes that all words in a sentence belong to only one topic, which is not in line with the actual situation. In fact, a sentence can also be a mixture of multiple topics. Our model samples a topic for each word instead of each sentence, which ensures that the sampled topic was more accurate. Besides, the most essential difference of our method is the difference of input compared with JST and ASUM. Their input are words of documents in corpora, while our inputs are aspect–opinion pairs of documents. In other words, we see each document as a bag of word pairs to better represent the corpora. The advantages of our method are as follows: firstly, the dimension of each document is greatly reduced, which greatly cuts the computation time of the model; secondly, compared with a single word, aspect–opinion pairs increase semantics to improve the accuracy of emotion classification; thirdly, it can output the associated linked network of each topic, not just word distribution, so as to facilitate human understanding of the results of emotion mining at the topic level.

In summary, the previous methods have been able to solve the problem of the emotional analysis for online reviews to a certain extent. But, because of neglecting the dependencies between words, these methods are more suitable for the case where the words in a text are independent of each other. In contrast, our model considers interword relationships and determines the number of topics beforehand to achieve the goal of reducing computation time, improving the accuracy rate of emotion mining at the topic level, and making the results easier to understand.

### 3. Proposed Model

In this section, we will show the framework of emotional mining for online reviews firstly, and then give a detailed description of the model we have proposed for online reviews. The model, named Topic-Specific Emotion Mining Model (TSEM), incorporates relationships between words and adds time stamps. Besides the model, we will also introduce the algorithms involved.

Figure 2 shows the framework of emotional mining for online reviews. At first, tremendous amounts of product review texts in one domain are collected. Then, we extract the comment content and time of each review. Step 2 extracts aspect–opinion pairs of each review by pattern matching based on part of speech. Step 3 establishes Topic-Specific Emotion Mining Model by extending LDA [10]. At step 4, the Gibbs sampling algorithm solves model parameters. Finally, the emotion distribution of each topic at different times can be output, which helps the user to understand each topic’s emotions of the product at a specific time. The associated linked network of positive and negative emotions under each topic can be output, which conveys more semantic information for human understanding.

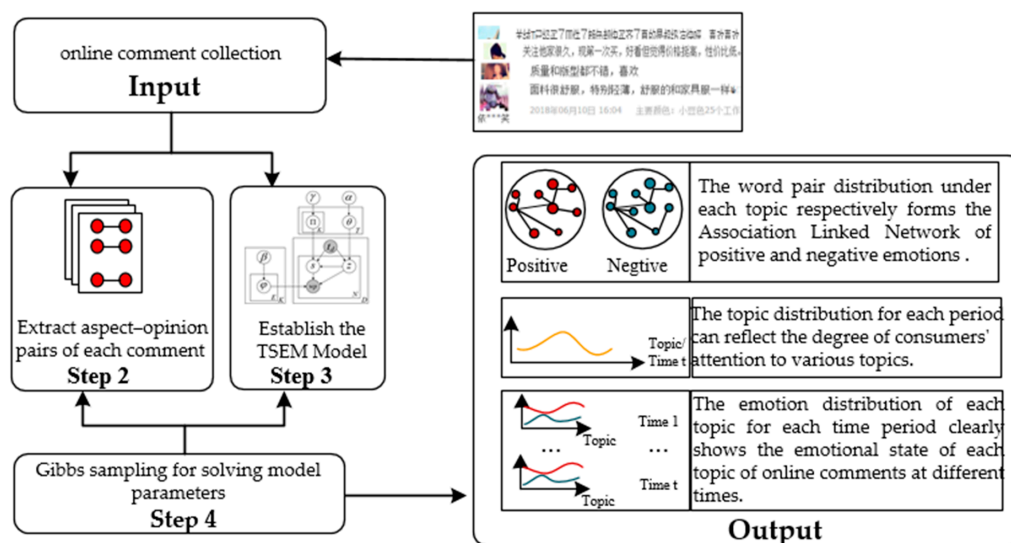
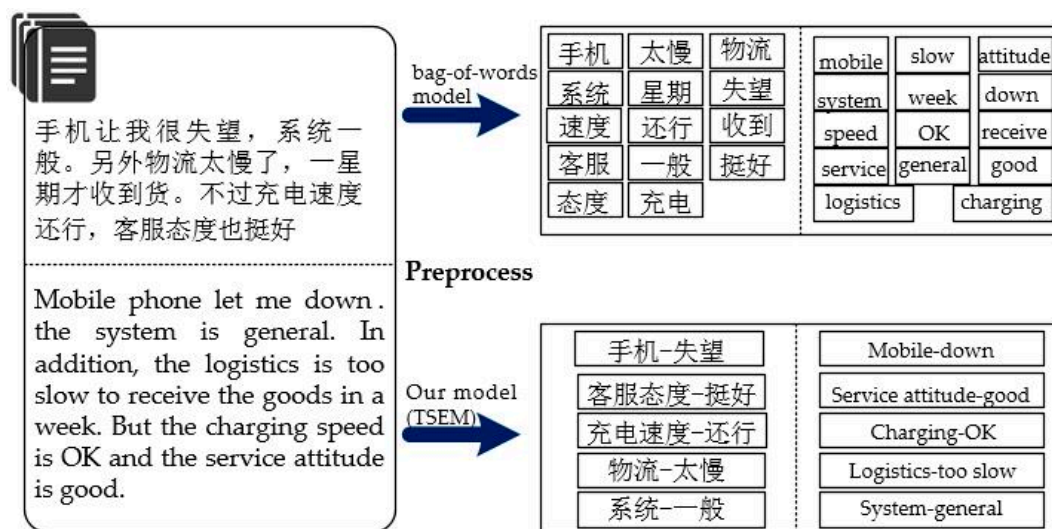


Figure 2. The overall framework of emotion mining for online comments.

### 3.1. Topic-Specific Emotion Mining Model

In this subsection, we mainly introduce Topic-Specific Emotion Mining model.

The emotion topic model, used to discover abstract topics and emotions in text, is a statistical model. It is also called the probability generation model, which is often used in natural language processing and machine learning. The traditional probabilistic generation model is the bag-of-words model. That is to say, the text is represented by a set of words in a bag, which has no interword order or syntactic structure. When a text contains views of different emotional types in many aspects, the word bag method cannot directly convey the text's viewpoint. TSEM is an extension of the traditional probability generation model. We believe that a subjective text is composed of the evaluation object and its corresponding evaluation content. There is a relationship between the evaluation object and the evaluation content, and the text is a set of opinions from different aspects. The different representations of text between a traditional bag-of-words model and the TSEM model are shown in Figure 3.



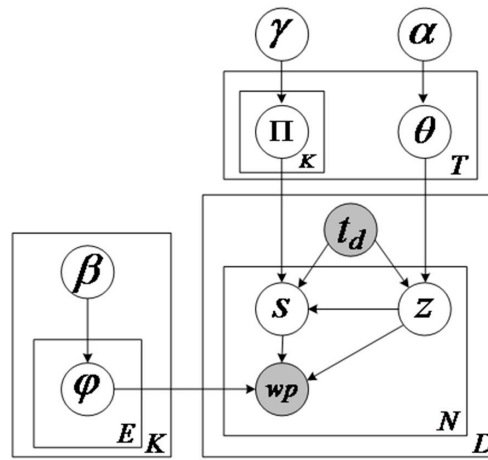
**Figure 3.** Representation of text by a traditional word bag model and the Topic-Specific Emotion Mining (TSEM) model.

The graphical model of TSEM for online comments is shown in Figure 4. Shadow nodes represent the observable data, the unshaded nodes indicate the hidden variables and the arrows mean dependencies. For comment text datasets, the comment time and words for each review are observable. Therefore, they are represented by shaded nodes in the graph model of TSEM. The explanation of notations involved in this paper is shown in Table 1. Usually, a buyer creates the emotion of a topic after determining the topic to be evaluated, instead of the other way around. Different from the traditional emotion-topic models, such as ASUM, emotion label sampling is affected by the topic in our model. Only when the topic and timestamp of a word pair are determined can the emotion of the word pair be generated. TSEM generates each aspect–opinion pair on the basis of the word pair's sentiment and the topic.

Our model, combining time, topic, and emotion, allows us to analyze emotional trends of various topics in different periods for reviews, besides grasping the time-specific topics of a product as a whole. Compared with other topic models, TSEM uses aspect–opinion pair distribution, which makes the result of emotion mining more understandable. In this paper, the collection of  $D$  review documents is formalized  $\{d_1, d_2, d_3, \dots, d_D\}$ . Each review document in the collection is composed of time  $t_d$  and a sequence of  $N_d$  word pairs, which can be defined as  $d = \{wp_1, wp_2, wp_3, \dots, wp_{N_d}, t_d\}$ . Again, we have built an aspect–opinion pair dictionary with  $V$  different items in the preprocessing, from which each word pair in each document is chosen and is denoted in its corresponding index. The value of  $wp_i$  ranges from 1 to  $V$ . Time of reviews is divided by 0.25 in this paper, and the number of time



windows is represented by  $T$ . Certainly, other time partition granularities can also be opted in line with different demands.



**Figure 4.** The graphical model of the Topic-Specific Emotion Mining (TSEM) model: unlike existing emotion topic models, emotion relies on topic rather than the opposite, and observable data ‘timestamps’ are added in our model.

**Table 1.** Description of notation used in this paper.

Notation	Description
$T$	Number of time-spans, e.g., quarter, in this paper
$D$	Number of comment documents
$N$	Number of words
$E$	Number of emotions
$V$	Size of vocabulary
$\alpha$	A $K$ -dimensional dirichlet prior vector for $\theta$
$\beta$	A $V$ -dimensional dirichlet prior vector for $\varphi$
$\gamma$	A $E$ -dimensional dirichlet prior vector for $\Pi$
$\theta_t$	A $K$ -dimensional topic distribution of $t$ th time-span
$\varphi_{ke}$	A $V$ -dimensional word pair distribution of the $k$ th topic for $e$ th sentiment
$\Pi_{tk}$	A $E$ -dimensional sentiment distribution of the $k$ th topic for $t$ th time-span
$t_d$	The time tag of the $d$ th document
$z_{dn}$	A topic assignment of the $n$ th word in the $d$ th document
$s_{dn}$	A sentiment assignment of the $n$ th word in the $d$ th document
$wp_{dn}$	The observed $n$ th aspect–opinion word pair in the $d$ th document
$s_{-i}$	A vector of sentiment assignment for all word pairs except the $i$ th word pair
$z_{-i}$	A vector of topic assignment for all words except the $i$ th word pair
$t_{-i}$	A vector of observed time-spans for all words except the $i$ th word pair
$wp_{-i}$	A vector of observed word pairs except the $i$ th word pair
$N_{kev}$	The number of the $v$ th unique words that are assigned to $e$ th sentiment and the $k$ th topic
$N_{tke}$	The number of words that are assigned to $e$ th sentiment and the $K$ th topic, within $t$ th time-span
$N_{tk}$	The number of words that are assigned to $k$ th topic, within $t$ th time-span

The number of topics and emotions are  $K$  and  $E$ , respectively. Because of the few neutral views, emotions are classified as positive and negative, ignoring the emotions of neutrality in our work.  $E$  is equal to 2. The generation of a review text  $d$  can be seen as the generation of each word pair in the review. There are three steps for producing an aspect–opinion pair. First of all, one topic label  $z$  is selected from time-specific topic distribution  $\theta_t$ , where time  $t_d$  is known. Then, an emotion label  $s$  is sampled from emotion distribution  $\Pi_{tk}$ , when time  $t_d$  and topic label  $z$  have been determined. At last, after the determination of topic  $z$  and emotion  $s$ , we randomly draw a word pair  $wp$  from the emotion–topic word pair distribution  $\varphi_{zs}$ .

The generative process of TSEM is described as follows:

1. Draw topic distribution for each time span:  $\theta_t \sim \text{Dirichlet}(\alpha)$ ;
2. Draw sentiment distribution for each pair of topic and time-span:  $\Pi_{tk} \sim \text{Dirichlet}(\gamma)$ ;
3. Draw aspect–opinion pair distribution for each pair of sentiment and topic:  $\varphi_{ke} \sim \text{Dirichlet}(\beta)$ ;
4. For each comment document  $d$ :
  - a. Choose a topic  $z \sim \text{multinomial}(\theta_{td})$ ;
  - b. Choose a sentiment  $e \sim \text{multinomial}(\Pi_{tk})$ ;
  - c. Generate an aspect–opinion pair  $wp \sim \text{multinomial}(\varphi_{ke})$ .

After generating the comment text set through the above steps, the parameters  $\theta$ ,  $\Pi$ , and  $\varphi$  are used to analyze emotions of the review texts.

### 3.2. Extraction of Aspect–Opinion Pairs

In this subsection, we will explain in detail how to extract aspect–opinion pairs from online reviews.

The traditional topic model does not take into account the relationship between words, resulting in semantic loss, which causes the emotion mining results to have low accuracy and be incomprehensible. To resolve this problem, we match opinions with the aspect words evaluated by them. In this way, each single word in the word bag is replaced by a word pair. In our paper, the method to extract an aspect–opinion pair is pattern matching based on part of speech.

To a large extent, the aspect word is a noun or a noun phrase, and the part of speech of an opinion is adjective, adverb, or verb. Here, nouns, adjectives, adverbs, and verbs are reserved for extracting aspect–opinion pairs. Adverbs are utilized to embellish adjectives or verbs. As a result, adverbs are merged with the part modified by them into a whole, and then the part of speech of the whole is annotated as the property of modifier, which not only simplifies the matching pattern, but also solves the problem of negation to some extent. When the opinion words are modified by negative words, it will lead to the emotional reversal of the opinion words. Because negatives are usually adverbs in Chinese, this problem can be solved by combining them with modifiers. For example, in the comment “the appearance is not good”, the phrase (not(adv.) good(adj.)) changes into a compound phrase (not good(adj.)). For the sake of brevity, nouns, adjectives, adverbs, and verbs are respectively abbreviated to n., a., d., and v. in the paper. The matching pattern we have designed is shown in Table 2. And the aspect–opinion pair extraction method based on pattern matching is shown in Algorithm 1.

---

#### Algorithm 1: Aspect–opinion pair extraction based on pattern matching

---

Input: A collection of user online comments  $D$

Output: Aspect–opinion pairs  $\langle Wa, Wo \rangle$  set  $S$  of each comment

1. for each comment text  $d$  in  $D$ :
  2.     Divide the comment text into comment fragments set based on separator (comma or space) in the text;
  3.     for each comment fragment in comment text  $d$ :
  4.         segment words and part-of-speech tagging using word segmentation tools;
  5.         reserve nouns, adjectives, verbs, and adverbs;
  6.         if adverb of existence in comment fragment:
  7.             merge adverbs with the part modified by them into a whole, and then annotate the part of speech of the whole as the property of modifier;
  8.         do pattern matching according to Table 2;
  9.         put aspect–opinion pairs  $\langle Wa, Wo \rangle$  into set  $S$ ;
  10.     end for
  11.     output the aspect–opinion pairs set  $S$ ;
  12. end for
-

**Table 2.** Simplified matching mode.

Matched Pattern	Examples
n.+a.	Taste (noun) is very (adv.) good (adj.)
n.+v.	Price (noun) increases (verb)
a.+n.	Lovely (adj.) package (noun)
v.+n.	Like (verb) screen(noun)
n.+a.+v.	Delivery (noun) fast (adj.) transport (verb)
n.+v.+a.	Sellers (noun) deliver (verb) fast (adj.)

In Chinese, people like to use commas or spaces to divide a comment into fragments. Each comment is divided into comment fragments according to the actual separator (comma or space) used by the comment. In order to reduce redundancy and improve accuracy, we performed pattern matching for each comment fragment instead of for the whole review.

Because of the differences between Chinese and English expressions, the matching pattern of aspect–opinion pairs is different. The matching pattern is designed for Chinese texts in this paper.

### 3.3. Gibbs Sampling for the Topic-Specific Emotion Mining Model

Topic-Specific Emotion Mining Model (TSEM), as an extension model of LDA [10], also applies Gibbs sampling algorithm [28] to the latent parameter estimation. Gibbs sampling will sequentially sample  $s_i$  and  $z_i$  from the distribution of variables given current values of all other variables and the data.

In our model, there are three types of distribution, which are  $\theta_t$ ,  $\Pi_{tk}$ , and  $\varphi_{ke}$ .  $\theta_t$  means topic distribution of the time  $t$ .  $\Pi_{tk}$  implies the time-specific and topic-specific emotion distribution.  $\varphi_{ke}$  refers to the aspect–opinion pair distribution under sentiment  $e$  and topic  $k$ . In order to obtain the three above-mentioned distributions, we firstly have to deduce a joint posterior distribution of emotion and topic for each word pair. The following is the description of the formula derivation result. The symbols in the formulas are defined in Table 1.

$$P(s_i = j, z_i = a | s_{-i}, z_{-i}, w, p_{-i}, t_{d-i}, w_i = m, t_i = y, \alpha, \beta, \gamma) \propto \frac{\{N_{ajm}\}_{-i} + \beta_m}{\{N_{aj}\}_{-i} + \sum_{v'=1}^V \beta_{v'}} \cdot \frac{\{N_{yaj}\}_{-i} + \gamma_j}{\{N_{ya}\}_{-i} + \sum_{e'=1}^E \gamma_{e'}} \cdot \frac{\{N_{ya}\}_{-i} + \alpha_a}{\{N_y\}_{-i} + \sum_{k'=1}^K \alpha_{k'}} \quad (1)$$

Formula (1) is the conditional posterior for  $s_i$  and  $z_i$ , which is the conditional probability derived by marginalizing out the random variables  $\theta_t$ ,  $\Pi_{tk}$ , and  $\varphi_{ke}$ . The subscript  $-i$  denotes a quantity that does not include data from  $i$ th position.

According to Formula (1), the Gibbs sampling algorithm respectively samples an emotional label and a topic label for each word pair of each comment. The sampling process is executed iteratively until the sample results are stable. As for the Gibbs sampling algorithm of our model, we make a generalization in Algorithm 2.

After getting the results of the sampling, we can use the sample obtained from the Markov chain to estimate the required posterior distribution. The three distributions can be estimated separately on the basis of Formulas (2)–(4).

The calculation formula of aspect–opinion pair distribution under sentiment  $e$  and topic  $k$  is:

$$\varphi_{kev} = \frac{N_{kev} + \beta_v}{N_{ke} + \sum_{v'=1}^V \beta_{v'}} \quad (2)$$



Formula (2) is the approximate estimated distribution of word pairs in topics and sentiment labels. Where  $N_{ke}$  is the number of word pairs assigned to the topic  $k$  and the emotional  $e$ .  $\beta_v$  means the  $v$ th element of the vector  $\beta$ .  $k$  is a topic token and  $e$  is an emotion token.

---

**Algorithm 2:** Gibbs sampling algorithm flow of TSEM
 

---

**Input:** Prior parameters  $\alpha, \beta, \gamma$ , the number of topic  $K$ , the number of sentiment  $E = 2$ , Time-words matrix  $W$

**Output:** Multinomial distribution  $\theta, \varphi, \Pi$ , label matrix  $S, Z$

1. Initialize topic label matrix  $Z$  and sentiment label matrix  $S$  randomly
  2. Calculate the value of  $K \times E \times V$  matrix  $\Omega$ ,  $T \times K \times E$  matrix  $\Psi$  and  $T \times K$  matrix  $\Phi$  based on matrix  $Z$  and matrix  $S$
  3. For iteration = 1 to Niter do
  4.   for  $i$  in  $T$  do
  5.   for  $j$  in  $\text{len}(\text{word}[i])$  do
  6.   a. Calculate the possibility  $P(s, z)$  based on Formula (1), which  $\text{word}[i][j]$  is assigned simultaneously to a sentiment label  $s$  and a topic label  $z$
  7.   a. Sample a sentiment label  $s$  based on Step 6
  8.   b. Sample a topic label  $z$  based on Step 6
  9.   c. Update the value of matrix  $\Omega$ , matrix  $\Psi$  and matrix  $\Phi$  based on the sample results of Step 7 and Step 8
  10.   end for
  11.   end for
  12. Calculate three distribution  $\theta_t, \Pi_{tk}$  and  $\varphi_{ke}$  based on Formula (2)–(4).
  13. End
- 

The estimation of topic distribution of the time  $t$  is:

$$\theta_{tk} = \frac{N_{tk} + \alpha_k}{N_t + \sum_{k'=1}^K \alpha_{k'}} \quad (3)$$

Formula (3) is the approximate estimated distribution over time for topics. Similarly,  $N_t$  is the number of word pairs that belong to the time  $t$ .  $\alpha_k$  means the  $k$ th element of the vector  $\alpha$ .  $t$  is a time token.

Finally, the prediction of time-specific  $t$  and topic-specific  $k$  emotion distribution can be formulated as:

$$\Pi_{tke} = \frac{N_{tke} + \gamma_e}{N_{tk} + \sum_{e'=1}^E \gamma_{e'}} \quad (4)$$

Formula (4) is the approximate estimated distribution of sentiment in topics and time.  $\gamma_e$  is the  $e$ th element of the vector  $\gamma$ . Other notations are described in Table 1.

## 4. Experiments

In this section, first of all, the dataset and preprocessing process are introduced. Secondly, we present the setup of prior parameters and how to determine the number of topics adaptively. Thirdly, we compare sampling time of different word frequencies with other models. Fourthly, the classification accuracy of aspect–opinion pairs under each topic is analyzed. Finally, we show the semantics of the output of our model.

### 4.1. Dataset and Preprocessing

We have gathered 27,157 online reviews of Yili Milk from Taobao (<https://www.taobao.com/>) from January 2017 to December 2017 as our dataset, generated by 19,916 users. On average, each comment contains seven comment fragments, and each comment fragment contains six words. Each review text consists of comment content and comment time. In the first place, we get rid of invalid

comments and sort the comments according to time. Then, we segment all words and part of speech tagging for each comment with Jieba Chinese text segmentation (<https://github.com/fxsjy/jieba>) and remove stop words. Our model only preserves nouns, verbs, adjectives, and adverbs. Last but not least, we use Algorithm 1 to further process reviews after segmenting words, since each review text is represented by a set of word pairs in our model.

Existing models that combine topic and emotion are the Joint Sentiment/Topic Model (JST) and the Aspect and Sentiment Unification Model (ASUM). Besides the TSEM's variant model TSEM\_n (TSEM with no relation), the two models mentioned above are also selected as comparison models.

In reality, not every topic has an emotional tag and cannot be manually annotated. However, it is clear whether the sentiment and topic of an aspect–opinion pair are classified correctly. We employ the evaluation index precision@n, where  $n$  is the  $n$ th position, to measure the accuracy of classification. It is worth noting that the accuracy of the classification includes both the accuracy of sentiment classification and the accuracy of topic classification.

In the online comments, some topics are dense and some topics are sparse, so the correct number of word pairs per topic is inconsistent. We determine  $n$  of precision@n with the following formula:

$$n_k = \frac{p_k}{p_{\max}} N_{\text{positive|negative}} \quad (5)$$

where  $p_k$  is the distribution probability of topic  $k$ ,  $p_{\max}$  is the maximum distribution probability in all topics, and here we set up  $N_{\text{positive}} = 30$  and  $N_{\text{negative}} = 20$ .

#### 4.2. Setting Parameters and Determining the Number of Topics

Our model involves the following parameters: hyper parameters  $\alpha$ ,  $\beta$ ,  $\gamma$ , the total number of topics  $K$ , and the total number of emotions  $E$ . Because neutral emotions are few, we ignore neutral emotions and divide emotions into two categories: positive and negative.  $E$  equals 2.  $\beta$  and  $\gamma$  are respectively set to 0.1 and 0.01 by using empirical values.

The number of topics is critical. Too many topics can lead to high similarity among topics, making it difficult to distinguish different topics. Too few topics will cause granularity of topics, meaning the model cannot calculate topic-level emotion. In a defined dataset, the number of topics should be already determined. Namely, the number of topics in a comment text dataset is certain but unknown. Perplexity [13] is a measure of the quality of the topic model. The smaller the perplexity value is, the better the topic model is. The formula for perplexity is as follows:

$$\text{perplexity} = \exp - \frac{\sum_{t=1}^T \ln(\prod_{k=1}^K \prod_{e=1}^E \prod_{v=1}^V (\varphi_{kev} * \pi_{tke} * \theta_{tk})^{n_t^{kev}})}{\sum_{t=1}^T n_t} \quad (6)$$

where  $n_t^{(kev)}$  is the number of word pairs  $v$  belonging to time  $t$  that are assigned to the topic  $k$  and the emotion  $e$ .

The value of parameter  $\alpha$  is commonly associated with the total number of topics  $K$ . Thus, we performed a set of experiments by adjusting the values of  $\alpha$  and  $K$  and calculated the perplexity of each experimental result. The experimental results are arranged as Figure 5.

In this experiment, the range of  $\alpha$  is 0.1 to 1.0, with step length of 0.1. As for the value of each alpha,  $K$  ranges from 5 to 50, for which the step length is 1. The experiment is iterated 2000 times for each pair of  $\alpha$  and  $k$ . Figure 5 shows the smallest perplexity in each value of  $\alpha$ . From Figure 5, it is easy to see that the minimum value of perplexity is obtained when  $\alpha$  is 0.6 and  $K$  is 19.

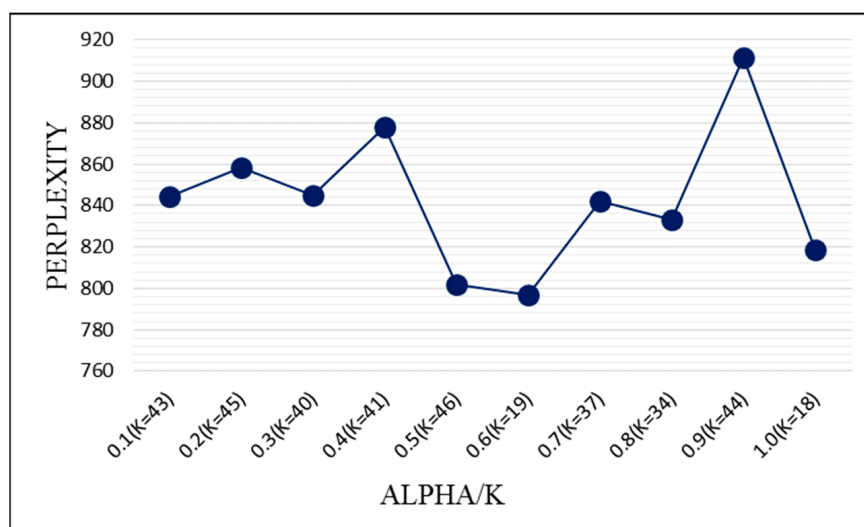


Figure 5. Perplexity with different values of  $\alpha$  and  $k$ .

In order to examine whether the number 19 is in line with the actual number of topics or not, we conducted a questionnaire survey on 20 students who frequently shop online to get aspects they pay close attention to when buying milk or yogurt. The results of the questionnaire are integrated to get the specific contents of the 19 topics. Table 3 shows the key words of 19 topics.

Table 3. Topics list.

topics	taste, price, logistics, package, crowd, date, customer service, courier, texture, quality, volume, cost performance, activity, whole, gift, delivery speed, distribution, quality guarantee period, nutrition
--------	--

#### 4.3. Comparison of Sampling Time with Existing Models

Retaining different word frequency affects the sampling time. To make it fair, our contrast models also retain nouns, adjectives, verbs, and adverbs in the text.

Figure 6 presents the computation time of TSEM, TSEM\_n, ASUM, and JST under different wfre, where wfre means the lower bound of word frequency of the words we retained. The wfre varies from 1 to 5. For instance, wfre = 1 means that we sample all the words. wfre = 2 means that we only sample words whose word frequency is greater than 1. Figure 6 shows the sampling time of 1000 iterations. As can be seen from Figure 6, the sampling time of TSEM under each word frequency is significantly lower than that of the other three methods. Our method implements dimensionality reduction representations of comment texts in accordance with the actual situation. When wfre = 1, the computation time of TSEM, TSEM\_n, ASUM, and JST is 3190 s, 8510 s, 5340 s, and 7150 s, respectively. This corresponds to sampling time reduction of 62.51%, 40.26%, and 55.38% relative to the other three models. This is because each input item is sampled as a topic and an emotion when solving the topic model. Using a bag of aspect–opinion pairs as input not only eliminates redundant words, but also greatly reduces the input items when an aspect word and an opinion word are combined. The fewer the input items, the fewer the total sampling times.

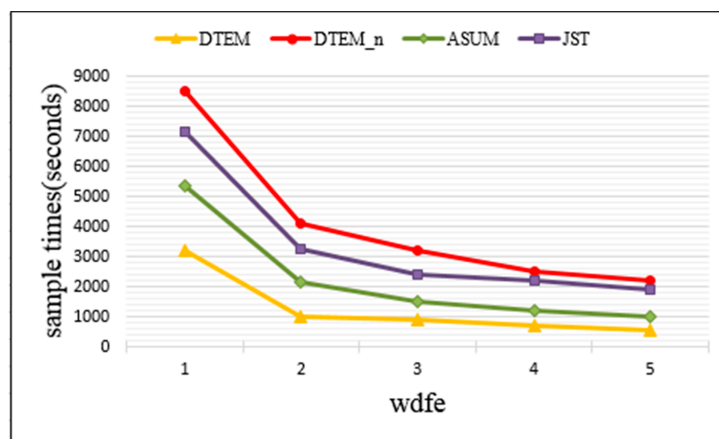


Figure 6. Occurrence of different words.

For each of the above methods, the calculation time drastically decreases when the value of wfre changes from 1 to 2. The reason is that the number of words or aspect–opinion pairs with word frequency of 1 is large. Thus, when the words or aspect–opinion pairs with word frequency of 1 are deleted, the input items decrease greatly, which causes the sampling time to decrease sharply. The trends of computing time tend to be mild when the value of wfre continues to increase. This is due to the relative decrease in the number of words or aspect–opinion pairs with word frequency greater than 1. Therefore, with the increase of the retained word frequency, the decline rate of the time curve slows down.

Experimental results prove that our model can substantially cut down the sampling time to speed up the operation.

#### 4.4. Evaluation of Classification Accuracy under Each Topic

For an established dataset, the aspects concerning consumers have been determined. Neither ASUM nor JST can automatically determine the number of product topics. The number of topics is set to 19 according to Section 4.1. The accuracy of each topic is evaluated below.

Figure 7 shows the accuracy of each topic's positive emotion, while the accuracy rate of each topic's negative emotion is shown in Figure 8. For the other three methods based on bag of words, the topic with the most related nouns in the top 50 words is the topic of the class. We then make manual judgments on the emotional classification and topic classification of the opinion words and count the number of opinion words correctly classified in each class. From Figures 7 and 8, we can see that different topics have different accuracy rates under the same emotions. This is due to the sampling bias caused by the lower discussion among users of some product topics. From Figures 7 and 8, the accuracy of the same topic varies under different emotions. Generally, the accuracy of positive emotions is slightly higher than that of negative emotions. This is because online reviews are generally more favorable than bad ones. Although the accuracy rate will change with topic, we can see from Figures 7 and 8 that our method is obviously higher than other methods.

The average accuracy of positive and negative emotions is shown as Figure 9, where the average accuracy of positive emotion of TSEM, TSEM\_n, ASUM and JST is 0.8567, 0.6951, 0.6269, and 0.5662, respectively, and that of negative emotion is 0.7617, 0.6416, 0.5634, and 0.5173. After further calculation, TSEM increased the average accuracy of the two emotions by 23.25%, 35.97%, and 49.37%, respectively, compared with the other three models. The reason is that the other three methods input words, but the topic and emotion of a word are often ambiguous, thus increasing the risk of inaccurate sampling results. The input of our method is aspect–opinion pairs.

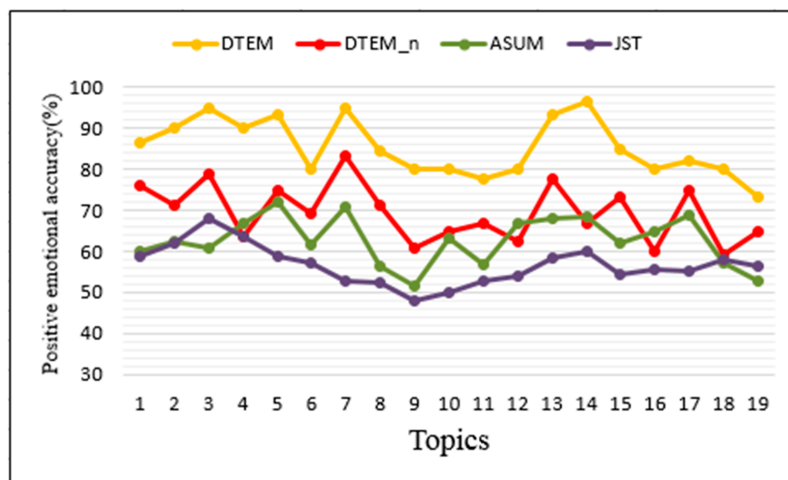


Figure 7. Positive accuracy for each topic.

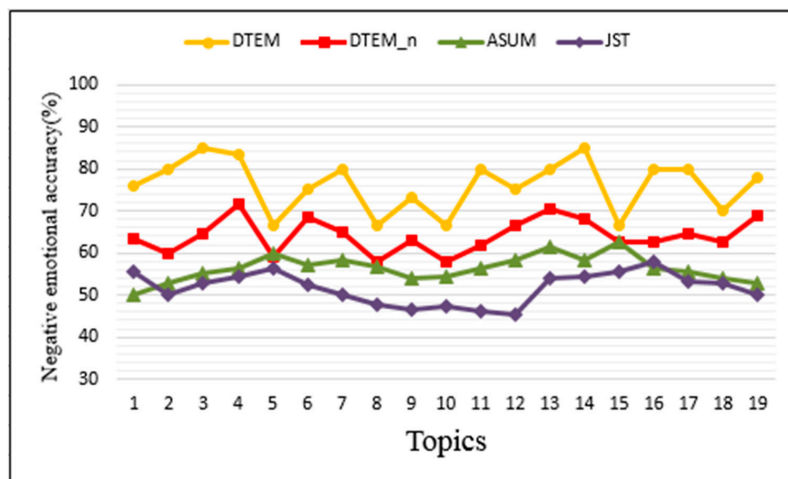


Figure 8. Negative accuracy for each topic.

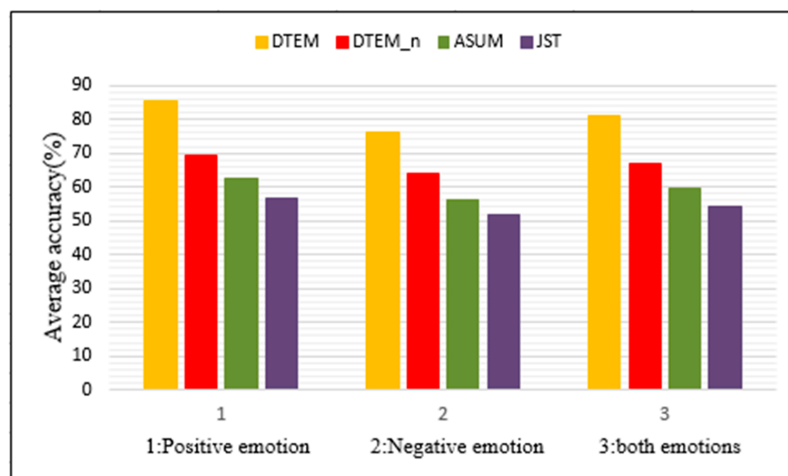


Figure 9. Average emotional accuracy of the four methods.

Through the analysis of experimental results, our model improves the emotional accuracy of the topic level, which is better than TSEM\_n, ASUM, and JST.



#### 4.5. Understandability of Results in Our Model

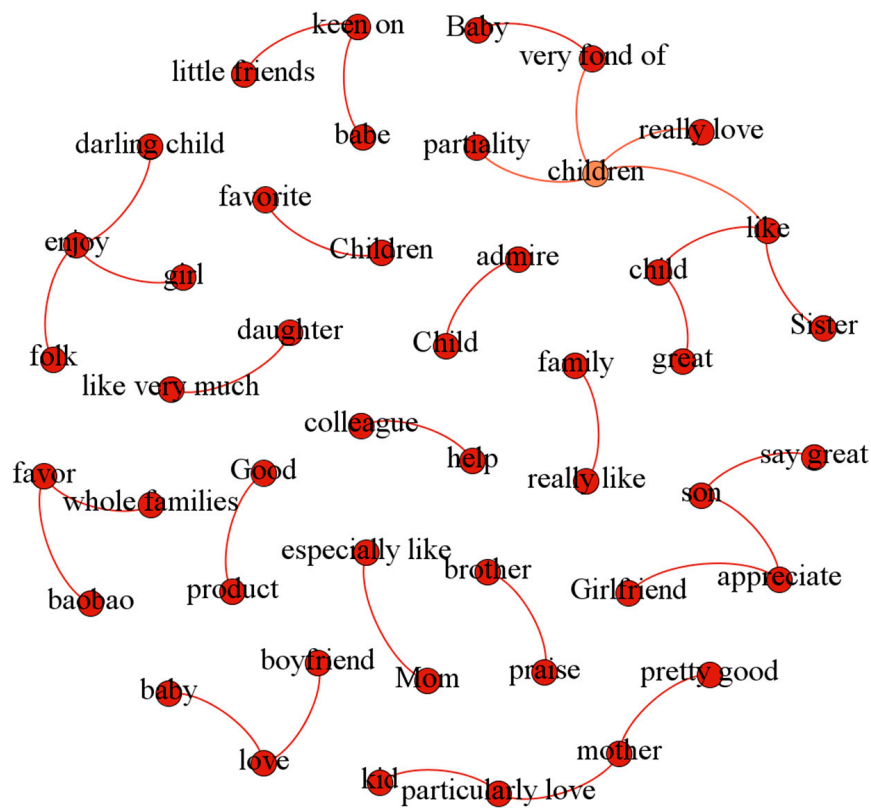
Our method uses aspect–opinion pairs instead of bag of words to represent review texts, which makes our outputs contain more semantic information than the other models. In our model output results, opinion words and aspect words are not mixed at random, but appear in the original collocation, which will undoubtedly help one understand the results.

The top 15 word pairs with their probability of the topic “Crowd’s positive emotion” and that of the topic “Package’s negative emotion” are enumerated in Table 4. Figures 10 and 11 are two associated networks, which are respectively generated from the top 31 word pairs with probability under the topic “Crowd’s positive emotion” and the topic “Package’s negative emotion”. Apart from aspect-specific opinion, such as ‘broken’, general opinion words do not reflect topic information. For example, the word ‘like’ cannot be judged as to which aspect it describes. Similarly, a topic word does not carry any emotional information, such as ‘child’, ‘brother’, and ‘box’. When the topic words and viewpoints are presented in the form of word pairs, such as “children-like”, “son-appreciate”, and “whole families-favor”, it can be seen clearly that the positive emotions of the topic ‘Crowd’ are reflected, indicating that the product is liked by all kinds of people. Word pairs like “packaging-breakage”, “Box-damaged”, and “wrapper-broken” not only reveal consumers’ dissatisfaction with the product packaging, but also show that the reason for the dissatisfaction is the damage of the packing box.

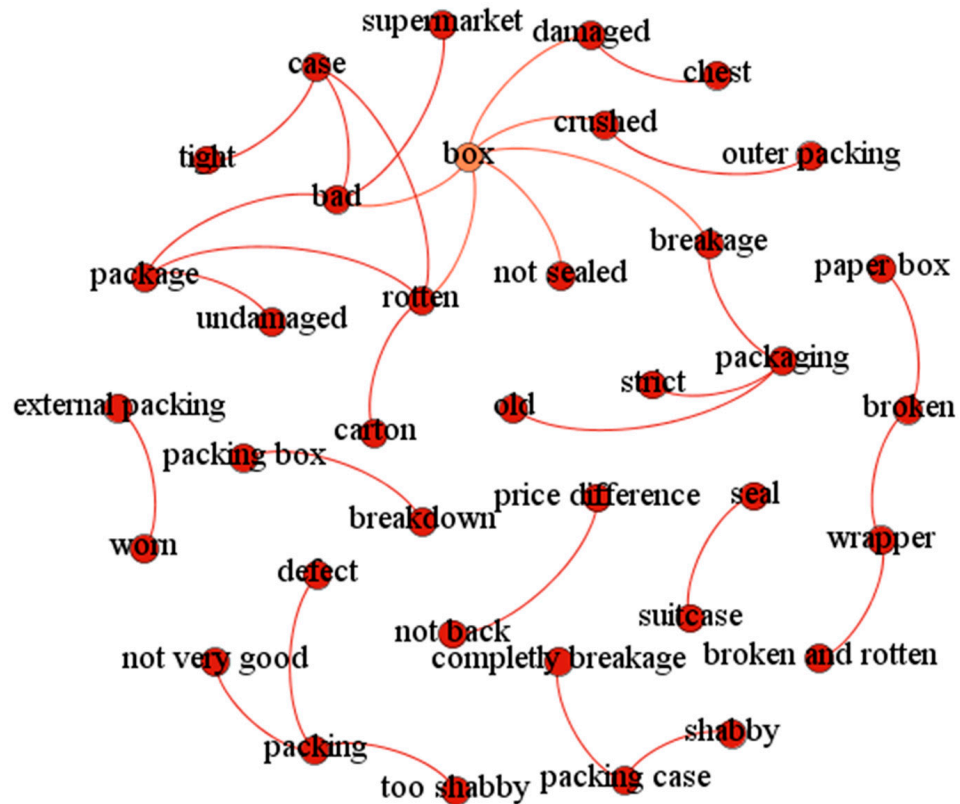
**Table 4.** Top 15 word pairs with probability of positive emotion of topic ‘Crowd’ and negative emotion of topic ‘Package’.

Crowd (Positive)			Package (Negative)		
Aspect–Opinion Pairs		Probability	Aspect–Opinion Pairs		Probability
children	like	0.156	packaging	breakage	0.041
baby	love	0.071	box	damaged	0.0265
son	appreciate	0.0526	wrapper	broken	0.008
darling child	enjoy	0.0385	package	bad	0.008
daughter	like very much	0.0217	packing box	breakdown	0.008
little friends	keen on	0.0195	package	rotten	0.0067
children	partiality	0.0098	case	rotten	0.0067
family	really like	0.0098	packing case	shabby	0.0054
child	great	0.0076	external packing	worn	0.0054
whole families	favor	0.006	chest	damaged	0.0054
brother	praise	0.0055	packaging	old	0.0054
children	very fond of	0.0055	case	tight	0.0054
child	like	0.0049	paper box	broken	0.0041
folk	enjoy	0.0049	wrapper	Broken and rotten	0.004
mother	particularly love	0.0044	box	breakage	0.004

To sum up, the TSEM model, in the form of word pairs, can exactly express more abundant semantic information, making the result of emotion mining easier for human understanding.



**Figure 10.** The association network under the positive emotion of the topic ‘Crowd’.



**Figure 11.** The association network under the negative emotion of the topic ‘Package’.

## 5. Conclusions

In recent years, online shopping platforms have accumulated massive numbers of online comments regarding products. Topic-level emotion mining of online comments helps companies and consumers to fully grasp the emotional state of purchasers on all aspects of products, which can effectively guide them to formulate marketing strategies and make consumption decisions. Different from the traditional topic model, this paper proposed a Topic-Specific Emotion Mining Model involving timestamps, aspects, and opinions, which considered the connections between words to better represent the comments. Our model, proven by the experimental results, is able to reduce computation time, improve the accuracy of topic-level emotional classification, and has highly semantic results for human understanding.

Our main conclusions can be stated as follows:

- (1) Since each comment was described as a bag of aspect–opinion pairs by including interword relationships, our model implemented the reduced-dimensional representation of comments, which greatly cut computation time.
- (2) The Topic-Specific Emotion Mining model constrains words' emotion and topic by adding interword relationships to improve the accuracy of sentiment classification at the topic level. Meanwhile, the number of topics in our model that are adaptively determined by calculating perplexity has also improved the emotion accuracy of the topic level.
- (3) In comparison with the output of the mainstream emotional topic models, our model containing rich semantic information is more interpretable.

The next step is to apply our model to the topic-level emotion mining for various online products. Certainly, our model is applicable to all kinds of data besides our experimental dataset. At present, our method is based on the characteristics of online comments and takes aspect–opinion pairs as the input of the model. The main application area is Chinese online comment emotion mining. The method we have proposed needs to be improved in the following two aspects. Firstly, optimization of the extraction method of aspect–opinion pairs is necessary, and secondly, the scope of application of the model needs to be extended. In the future, we can consider the dependency relationship between words in sentences and use dependency parsing to optimize the extraction of aspect–opinion pairs. For texts in other fields, we should specifically analyze text features and incorporate appropriate interword relations in order to obtain the most accurate results of emotional mining in more fields. By doing so, we can extend the application scope of the model.

**Author Contributions:** Conceptualization, Y.Y.; Methodology, Y.Y.; Supervision, X.L.; Writing—original draft, Y.Y.

**Funding:** The research work reported in this paper was supported in part by the National Science Foundation of China under grant NO. 91746203, 61471232 and 61625304. This work was jointly supported by the Shanghai Science International Cooperation Project under grant NO.16550720400.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Svobodová, L.; Hedvičáková, M. Actual Situation and Development in Online Shopping in the Czech Republic, Visegrad Group and EU-28. In *Modern Approaches for Intelligent Information and Database Systems*; Springer: Cham, Switzerland, 2018; Volume 769, pp. 269–279.
2. Saumya, S.; Singh, J.P.; Baabdullah, A.M.; Rana, N.P.; Dwivedi, Y.K. Ranking Online Consumer Reviews. *Electron. Commer. Res. Appl.* **2018**, *29*, 78–89. [[CrossRef](#)]
3. Yap, C.S.; Ong, M.Y.; Ahmad, R. Online Product Review, Product Knowledge, Attitude, and Online Purchase Behavior. *Int. J. E-Bus. Res.* **2017**, *13*, 1–19. [[CrossRef](#)]
4. Tripathy, A.; Agrawal, A.; Rath, S.K. Classification of Sentiment Reviews using n-Gram Machine Learning Approach. *Expert Syst. Appl.* **2016**, *57*, 117–126. [[CrossRef](#)]

5. Wan, Y.; Nie, H.; Lan, T.; Wang, Z. Fine-grained Sentiment Analysis of Online Reviews. In Proceedings of the 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Zhangjiajie, China, 15–17 August 2015; pp. 1406–1411.
6. Zhang, H.; Rao, H.; Feng, J. Product Innovation Based on online Review Data Mining: A Case Study of Huawei Phones. *Electron. Commer. Res.* **2018**, *18*, 3–22. [[CrossRef](#)]
7. Ganu, G. Improving the Quality of Predictions using Textual Information in Online User Reviews. *Inf. Syst.* **2013**, *38*, 1–15. [[CrossRef](#)]
8. Dang, Y.; Zhang, Y.; Chen, H.C. A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews. *IEEE Intell. Syst.* **2010**, *25*, 46–53. [[CrossRef](#)]
9. Ye, Q.; Zhang, Z.; Law, R. Sentiment Classification of Online Reviews to Travel Destinations by Supervised Machine Learning Approaches. *Expert Syst. Appl.* **2009**, *36*, 6527–6535. [[CrossRef](#)]
10. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
11. Lin, C.; He, Y. Joint Sentiment/Topic Model for Sentiment Analysis. In Proceedings of the 18th ACM Conference on Information and Knowledge Management, Hong Kong, China, 2–6 November 2009; pp. 375–384.
12. Jo, Y.; Oh, A.H. Aspect and Sentiment Unification Model for Online Review Analysis. In Proceedings of the 4th ACM International Conference on Web Search and Data Mining, Hong Kong, China, 9–12 February 2011; pp. 815–824.
13. Azzopardi, L.; Girolami, M.; Risjbergen, K.V. Investigating the Relationship between Language Model Perplexity and IR Precision-recall Measures. In Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval, Toronto, ON, Canada, 28 July–1 August 2003; pp. 369–370.
14. Mulkalwar, A.; Kelkar, K. Sentiment Analysis on Movie Reviews Based on Combined Approach. *Int. J. Sci. Res.* **2012**, *3*, 1739–1742.
15. Augustyniak, L.; Kajdanowicz, T.; Kazienko, P.; Kulisiewicz, M.; Tuliglowicz, W. An Approach to Sentiment Analysis of Movie Reviews: Lexicon Based vs. Classification. In Proceedings of the 9th International Conference on Hybrid Artificial Intelligence Systems, Salamanca, Spain, 11–13 June 2014; pp. 168–178.
16. Jararweh, Y.; Al-Ayyoub, M.; Smadi, M.A.; Obaidat, I.; Mohawesh, R. Using Enhanced Lexicon-Based Approaches for the Determination of Aspect Categories and Their Polarities in Arabic Reviews. *Int. J. Inf. Technol. Web Eng.* **2016**, *11*, 15–31.
17. Kennedy, A.; Inkpen, D. Sentiment Classification of Movie Reviews Using Contextual Valence Shifters. *Comput. Intell.* **2006**, *22*, 110–125. [[CrossRef](#)]
18. Prabowo, R.; Thelwall, M. Sentiment Analysis: A Combined Approach. *J. Informetr.* **2009**, *3*, 143–157. [[CrossRef](#)]
19. Zhu, J.; Wang, H.; Zhu, M.; Tsou, B.K.; Ma, M. Aspect-Based Opinion Polling from Customer Reviews. *IEEE Trans. Affect. Comput.* **2011**, *2*, 37–49.
20. Poria, S.; Cambria, E.; Ku, L.-W.; Gui, C.; Gelbukh, A. A Rule-based Approach to Aspect Extraction from Product Reviews. In Proceedings of the Second Workshop on Natural Language Processing for Social Media, Dublin, Ireland, 24 August 2014; pp. 28–37.
21. Schouten, K.; Frasinicar, F. Finding Implicit Features in Consumer Reviews for Sentiment Analysis. In Proceedings of the International Conference on Web Engineering, Toulouse, France, 1–4 July 2014; pp. 130–144.
22. Jeyapriya, A.; Selvi, C.K. Extracting Aspects and Mining Opinions in Product Reviews using Supervised Learning Algorithm. In Proceedings of the 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Coimbatore, India, 26–27 February 2015; pp. 548–552.
23. Bhoir, P.; Kolte, S. Sentiment Analysis of Movie Reviews using Lexicon Approach. In Proceedings of the 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, India, 10–12 December 2016; pp. 1–6.
24. Mei, Q.; Ling, X.; Wondra, M.; Su, H.; Zhai, C. Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs. In Proceedings of the 16th International Conference on World Wide Web, Banff, AB, Canada, 8–12 May 2007; pp. 171–180.
25. Bagheri, A.; Saraee, M.; Jong, F.D. ADM-LDA: An Aspect Detection Model Based on Topic Modelling using the Structure of Review Sentences. *J. Inf. Sci.* **2014**, *40*, 621–636. [[CrossRef](#)]

26. Wang, S.; Chen, Z.; Liu, B. Mining Aspect-Specific Opinion using a Holistic Lifelong Topic Model. In Proceedings of the 25th International Conference on World Wide Web, Montréal, QC, Canada, 11–15 April 2016; pp. 167–176.
27. Park, D.H.; Zhai, C.X.; Guo, L. SpecLDA: Modeling Product Reviews and Specifications to Generate Augmented Specifications. In Proceedings of the 2015 SIAM International Conference on Data Mining, Vancouver, BC, Canada; 2015; pp. 837–846.
28. Griffiths, T. Gibbs Sampling in the generative model of Latent Dirichlet Allocation. Available online: <https://people.cs.umass.edu/~wallach/courses/s11/cmptsci791ss/readings/griffiths02gibbs.pdf> (accessed on 24 March 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).