Journal of Computers Vol. 32 No. 3, 2021, pp. 43-56
doi:10.3966/199115992021063203004

Thought Expression Analysis of English Essays Combining Opinion and Sentiment

Yabing Wang, Guimin Huang*, Xiaowei Zhang, Yiqun Li, Maolin Li, Hui Li, Jun Li

Guangxi Key Laboratory of Image and Graphic Intelligent Processing, School of Computer Science and Information Security, Guilin University of Electronic Technology Guilin 541004, China
bulingyouzi@163.com, sendhuang@126.com

Received 15 July 2020; Revised 2 November 2020; Accepted 30 December 2020

Abstract. With the rapid development of science, technology and the gradual information of education, the automatic English essays correction system has received more and more attention. The existing automatic English essays correction system can provide feedback on vocabulary and grammar, but it has a lack of thought expression analysis of English essays. Therefore, it is impossible to comprehensively evaluate whether opinion of English essays is relevant and expression of sentiment is clear. To solve this problem, we propose to analyze the thought expression of English essays by combining sentiment analysis and opinion analysis. Through experiment analysis and verification from the aspects of relevance degree of opinion, sentiment tendency, and thought expression of English essays, the results show the proposed method has a good performance in analyzing thought expression of English essays.

Keywords: thought expression, opinion and sentiment, SP-LDA

1 Introduction

In recent years, with the continuous advancement of artificial intelligence and natural language processing technology, the teaching field has become more and more technological. The automatic correction system of English essays uses natural language processing, machine learning and other related technologies to automatically correct the words, grammar, and coherence of English essays, and gives scores and comments [1]. In English writing training, it can not only give English learners timely feedback and suggestions, and reduce the cost of manual correction in English tests, but also help to improve the enthusiasm of English learners to learn English [2]. However, after a period of practice, we find that the existing correction system can provide feedback on vocabulary and grammar, but lack analysis of thought expression of English essays. The thought expression of English essays includes two parts in scoring criteria: analysis of opinion and sentiment. As a result of this, there is a shortcoming that it is impossible to comprehensively evaluate whether the opinion of English essay is relevant and the expression of sentiment is clear. Starting from the internal quality of English essays, this paper analyzes the relevance degree of opinion and sentiment tendency of English essays, and realizes the function of analyzing thought expression of English essays to improve the function of the automatic correction system and its validity.

In terms of opinion analysis of English essays, the classical theme method Latent Dirichlet Allocation (LDA) [3] can transforms textual content information into digital information that is easy to model, which has a significant effect on mining potential themes of texts. But there are still some shortcomings. The generation method of LDA based on the assumption of "bag of word (BOW)", which does not take into account the location information and context information of words, resulting in the loss of text semantic information [4]. Huang G. [5] et al. added the sentence level analysis based on LDA and proposed SLDA algorithm applied in the automatic error correction system with good effect. However, SLDA still has the problem of generating too many themes, resulting in a low accuracy of final extracted

* Corresponding Author

theme. Therefore, this paper adds a paragraph-level theme based on SLDA, and proposes an algorithm——SP-LDA (Sentences-Paragraph Latent Dirichlet Allocation), which combines text chapter paragraph structure to improve the accuracy of theme extraction.

In terms of sentiment analysis of English essays, the sentiment analysis method based on sentiment lexicons is unsupervised method, which mainly relies on sentiment lexicons to calculate the sentiment tendency of essays [6]. The traditional method based on sentiment lexicons directly added the sentiment scores of words in sentences, but such simple accumulation ignores the context information of words, resulting in the low accuracy of sentiment analysis. For example, for the sentence, "This is not a good product.", "good" indicates positive sentiment polarity, but the sentiment shifter "not" changes the sentiment polarity of "good" in the sentence. For another sentence, "She is a very pretty girl.", "pretty" indicates positive sentiment polarity, and the increase word "very" enhances the intensity of "pretty". These problems are the traditional sentiment analysis methods based on sentiment lexicons can't deal with. To solve the above problems, we consider the influence of sentiment shifter, sentiment comparison, dynamic modality, coordinating conjunction, increase and decrease word on sentiment words. We judge the sentiment tendency of English essays from word level, sentence level and text level by connecting the context information of words.

The main contributions of this paper are as follows:

(1) Proposing SP-LDA by adding the paragraph-level theme, which combines text chapter paragraph structure to improve the accuracy of theme extraction.

(2) Considering the influence of sentiment shifter, sentiment comparison, dynamic modality, coordinating conjunction, increase and decrease word on sentiment analysis. Proposing the sentiment calculation methods of words, sentences and texts. Judging the sentiment tendency of English essays from word level, sentence level and text level by connecting the context information to solve the problem of low accuracy of sentiment analysis.

(3) Applying the opinion analysis technique and sentiment analysis technique to the automatic English essays correction system to solve the problem of lacking of thought expression analysis of English essays.

(4) Through experiment analysis and verification from the aspects of opinion analysis, sentiment analysis and thought expression analysis of English essays, the results show the proposed method has a good performance in analyzing the thought expression of English essays.

The remainder of this paper as follows. Section 2 presents the related work of opinion analysis and sentiment analysis of English essays; Section 3 describes SP-LDA and sentiment analysis method proposed in this paper in detail; Section 4 contrasts and analyses the experiment results; Section 5 summarizes the work of this paper and looks forward to the next work.

2 Related Work

In recent years, opinion analysis and sentiment analysis have received a lot of attention from industry and academia.

LDA is a semantic analysis method based on probabilistic. It is mainly used to mine the theme of texts. It has been widely studied and applied since its proposed [7], but it lost text semantic information because of not considering the location information and context information of words. Xu, G. [8] et al. embedded an extended vocabulary of sensitive words in LDA, and proposed a theme recognition method of the network sensitive information based on sensitive word weighted-LDA model to solve the problem of low recognition rate of sensitive information in existing theme recognition methods, while the proposed model has a large dependence on the sensitive vocabulary. Bastani, K. [9] et al. proposed an intelligent method based on LDA to extract the potential themes of consumer complaints and explore their changing trends over time, but the method is not very efficient. Gao, Y. [10] et al. proposed a combined LDA method -- CLDA and an online CLDA method to improve the accuracy of user interest inference, while the method still performs not well. Al. Helal, et. al., [11] used the combination of LDA and bigram to extract the potential themes of Bengali news, while the method is not accurate enough because it does not consider more influential factors for opinion analysis. Porter, K. et al., [12] used LDA to analyze public forums such as "DarkNet Markets" and reveal keywords and trends related to criminal activities, while the accuracy of this method is low. Huang, C. [13] et al. proposed a season theme method STLDA based on LDA to capture the theme of each scenic spot in different seasons to enrich tourism recommendation, while this method cannot be applied to other domains. Nabli, H. [14] et al.

proposed an adaptive semantic-focused crawler algorithm based on LDA for efficient cloud service discovery, while it can be spoofed by adding relevant words in a document. Huang, G. [5] et al. added sentence level analysis based on LDA and proposed SLDA algorithm to applied in the automatic error correction system with good results, while it generates too many themes.

The lexicon-based sentiment analysis method obtains the sentiment tendency of texts by calculating the sentiment word score. Khoo C. S. G. [15] et al. introduced a new general sentiment Lexicon WKWSCI, and compared it with five common sentiment Lexicon. The results showed that Hu & Liu Opinion Lexicon is best for product review texts, and WKWSCI Lexicon is best for non-review texts, while WKWSCI contains limited vocabulary and does not provide sufficient information. Trinh, S. [16] et al. constructed a Vietnamese sentiment lexicon to analyze the sentiment of Vietnamese Facebook data, while the sentiment lexicon does not perform well in other areas. Thompson, J. J. [17] et al. extended a lexicon-based sentiment extractor SO-CAL to analyze the sentiment information of users in 1000 games of StarCraft 2, while it is not clear whether this method can be applied to other fields. Khan, F. H. [18] et al. proposed a semi-supervised sentiment analysis method combining lexicons with the machine learning method, and corrected the sentiment score by using mathematical methods such as information gain and cosine similarity to solve the problem of unavailability and sparsity of data, while this method is not scalable. Dhaoui, C. [19] et al. conducted automatic sentiment analysis on consumer-generated content (CGC) by adopting Lexicon-based method, machine-learning method and the combination of the two respectively, the results showed that the accuracy of the two methods was similar, and the combination of the two methods significantly improved the classification performance of positive sentiment, while it still didn't do well in negative sentiment. Asghar, M. Z. [20] et al. considered emoticons, modifiers and domain specific terms on the basis of Lexicon-enhanced sentiment analysis which based on Rule-based classification scheme, and improved the performance of online comment sentiment analysis, while this method requires a classification system to classify and score words in specific domain. Saif, H. [21] et al. proposed a Lexicon-based method-SentiCircles to analyze sentiment on Twitter, which considered the co-occurrence patterns of words in different contexts to capture their semantics and update their intensity and polarity in sentiment lexicons accordingly, and it can analyze sentiment at both entity-level and tweet-level, while SentiCircles is not accurate enough because of not considering more influential factors for sentiment analysis.

In the above research works, most of the research objects are blog texts, product reviews, microblog texts, rather than English essays. This paper applies the opinion analysis technique and sentiment analysis technique to the automatic English essays correction system to solve the problem of lacking thought expression analysis of English essays.

3 Methodology

3.1 Opinion Analysis Method

3.1.1 Sentences Latent Dirichlet Allocation (SLDA)

SLDA adds a sentence-level theme based on the LDA. The text generation process of SLDA is based on the assumption that "a sentence only expresses one theme" and takes a sentence instead of a word as the processing unit. In the SLDA method, assuming there are *T* potential themes in a text, the probability formula of the *jth* word w_{ij} in the *ith* sentence of the text is shown in (1):

$$p(w_{i,j}) = \sum_{z=1}^{T} p(w_{i,j} \mid s_i = z, t_j = z) p(s_i \mid t_j = z) p(t_j = z). \qquad \textbf{(1)}$$

t_j represents the themes to which the word w_{ij} belongs, $p(t_j = z)$ represents the probability of occurrence of themes z in the text, $p(s_i \mid t_j = z)$ represents the probability that *ith* sentence in the text belongs to themes z, $p(w_{i,j} \mid s_i = z, t_j = z)$ represents the probability that the word w_{ij} belongs to themes z under the condition that the themes of the sentence s_i is z, and T represents the number of themes in the text.

3.1.2 Sentences Paragraph-Latent Dirichlet Allocation (SP-LDA)

SP-LDA is different from the three-layer structure of LDA and the four-layer structure of SLDA. It is a five-layer structure which contains words, sentences, themes, paragraphs, and texts. By adding the paragraph-level theme based on SLDA, SP-LDA combines text chapter paragraph structure to improve the accuracy of theme extraction. The structure of SP-LDA is shown in Fig. 1.

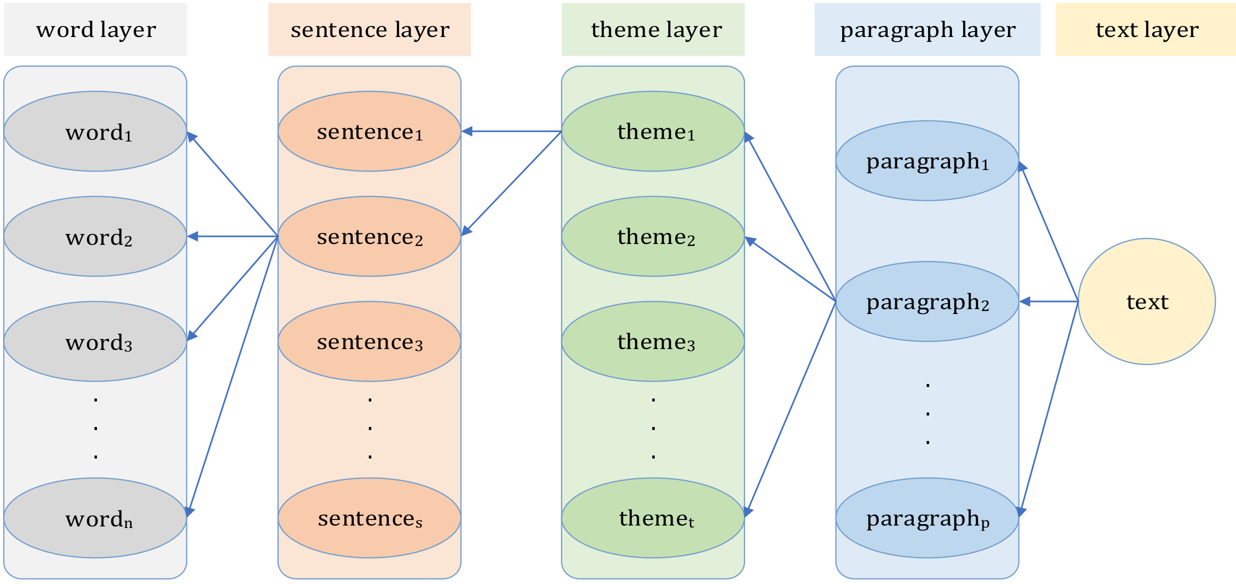

Fig. 1. The structure of SP-LDA

We added a paragraph theme based on SLDA in Fig. 1. In general, a sentence expresses a theme, while the themes of sentences in a paragraph are always similar. After getting the theme-word distribution, theme-sentences distribution, we get the paragraph-theme matrix distribution by the times of each theme selects the sentence in the paragraph, and then get the maximum theme of the paragraph. The sentences that corresponding to the maximum theme of the paragraph are grouped into the same paragraph according to the theme-sentence matrix. The probability method diagram of SP-LDA is shown in Fig. 2.

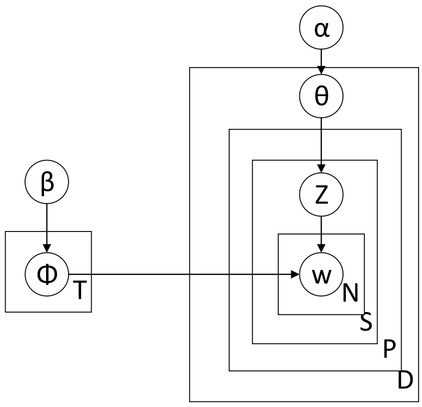

Fig. 2. The probability method diagram of SP-LDA

As shown in Fig. 2, α and β are hyper parameters of text-theme distribution θ and theme-word distribution ϕ respectively, T represents the number of themes, D represents the number of texts, P represents the number of paragraphs, S represents the number of sentences, and N represents the number of words w in the sentence, Z represents the potential theme variable.

If there is a text set with D texts and T themes, the processing process with SP-LDA is as follows:

(1) For each theme t, polynomial distribution ϕ_t is extracted from $Dirichlet(\beta)$ to get theme-word polynomial distribution;

(2) For each text d in the text set, the polynomial distribution θ_d is extracted from $Dirichlet(\alpha)$ to get text -theme polynomial distribution;

(3) For each sentence in text d, follow the steps below until all texts in the text set are processed:

a. To extract themes $T_{d,s}$ from the polynomial distribution $Multinomial(\theta_d)$;

b. To generate the words $W_{d,s}$ under the theme from the polynomial distribution $Multinomial(\phi_t)$ according to the extracted theme.

(4) To get the maximum theme of the paragraph according to the paragraph-theme matrix, and then the sentences that corresponding to the maximum theme of the paragraph are grouped into the same paragraph according to the theme-sentence matrix.

The pseudo-code of SP-LDA algorithm is shown in Algorithm 1:

Algorithm 1. Pseudo-code Of SP-LDA Algorithm

```
INPUT: parameters α, β, number of themes T, number of text D, number
of paragraphs P, number of sentences S, number of words N
OUTPUT: theme-word matrix distribution, text-theme matrix
distribution, paragraphs-theme matrix distribution, theme-sentences
matrix distribution
  BEGIN
      For (int t=0; t<T)
         Theme Word Matrix=draw (Dirichlet(β));
      For (int d=0; d<D)
         Text Theme Matrix=draw (Dirichlet(α));
           For (int p=0; p<P)
             For (int s=0; s<S)
               To choose (text theme Matrix);
               For (int n=0; n<N)
                To calculate (theme Word Matrix);
     To return Theme Word Matrix, Text Theme Matrix;
  END
```

In the process of iterative solution of algorithm 1, the probability of each word w in English essays assigned to each theme t shown in (2):

$$P_{word\ w\ assigned\ to\ theme\ t} = \prod_{i=1}^{n} \frac{\beta + N_1 + i}{\beta \times N_2 + N_3 + i}. \tag{2}$$

In formula (2), β =1, N_1 is the times of theme t selects word w, N_2 is the total number of words in English essay, N_3 is the total times of theme t is assigned, t = 1, 2, 3, ..., T, T is the number of theme set in English essays, w = 1, 2, 3, ..., N_2, i = 1, 2, 3, ..., n, n is the total times of occurrences of word w in English essay.

The probability of each sentence s in English essays assigned to each theme t shown in (3):

$$P_{sentence\ s\ assigned\ to\ theme\ t} = \frac{\alpha + N_4}{\alpha \times T + N_5} \times \prod_{w=1}^{N} P_{word\ w\ assigned\ to\ theme\ t}. \tag{3}$$

In formula (3), α =1, N_4 is the times of theme t selects sentence s, N_5 is the total number of sentences in English essay, s = 1, 2, 3, ..., N_5, w = 1, 2, 3, ..., N. The paragraph-theme matrix is obtained by the times of each theme selects the sentence in the paragraph, and then obtained the maximum theme of the paragraph from it.

Through Algorithm 1, we obtain the text-theme distribution and theme-word distribution of English essays. We assume that an English essay has only one maximum theme, and the maximum theme of English essays can be obtained from the text-theme distribution. Then find the opinion words corresponding to the maximum theme from the theme-word distribution. Finally, we use the text similarity algorithm [22] to calculate the similarity between the theme of English essays to be analyzed and the model English essays, which is the opinion score of English essays to be analyzed. The calculation formula of theme similarity is shown in (4):

$$Sim = \frac{\sum_{d=1}^{D}\left(\sqrt{\sum_{t=1}^{T}\left(P_1 - P_2\right)^2}\right)}{D}. \tag{4}$$

In formula (4), P_1 is the probability vectors of the theme of model English essays d is t, P_2 is the probability vectors of the theme of English essays to be analyzed is t, d = 1, 2, ..., D.

3.2 Sentiment Analysis Method

The key point of sentiment analysis method based on sentiment Lexicons is how to find sentiment words and calculate sentiment tendency of English essays. Adjectives, verbs, adverbs and nouns often represent the subjective feelings of the author. Therefore, this paper extracts these words as sentiment words to calculate the sentiment score of sentence and text using the sentiment Lexicons.

3.2.1 Context Features of Sentiment Words

On the basis of traditional sentiment analysis methods based on sentiment lexicons, we consider the influence of contextual information of sentiment words in the sentence. The contextual features of sentiment words considered in this paper include sentiment shifter, sentiment comparison, dynamic modality, coordinating conjunction “but”, increase and decrease word.

Coordinating conjunction “but”: “but” is usually used to join two opposing sentences. Therefore, “but” usually changes the sentiment polarity of the whole sentence [23].

Sentiment shifter: sentiment shifter includes presupposed words such as “hardly”, “barely”, “seldom”, and negator such as “not”, “never”, “none”, etc. Sentiment shifter generally changes the sentiment polarity of words in the sentence. We use coordinating conjunction “but” and sentiment shifter uniformly.

Sentiment comparison: sentiment comparison does not change the sentiment polarity of words, more and larger generally increase the intensity of words, less and small generally decrease the intensity of words.

Dynamic modality: dynamic modality has a great influence on the expression of sentiment in sentences. In this paper, we think that the sentiment intensity of “may”, “can”, “should”, “must” gradually increases.

Increase and decrease words: expressions of increase and decrease words do not change the polarity of words, only increase and decrease its intensity. We use sentiment comparison, dynamic modality, increase and decrease word uniformly.

3.2.2 Judgment of Sentiment Tendency of English Essays

(1) The Calculation of Sentiment Words Weight

We extract adjectives, nouns, verbs, and adverbs in English essays as sentiment words, then divide the words into strong positive, positive, weakly positive, weakly negative, negative and strong negative sentiment tendencies according to the sentiment scores of the words. We give different weight according to the different sentiment tendency of words, among which the corresponding weights of strong positive, positive, weakly positive, weakly negative, negative and strong negative sentiment tendency are respectively 3, 2, 1, - 1, - 2, - 3.

(2) The Calculation of Sentence Sentiment Score

We have obtained the weight of sentiment words in English essays from the above steps. We can calculate the sentiment score of each sentence in English essays by extracting the contextual features that have an impact on the sentiment words. Its calculation formula is shown in (5):

$$Sentence\ score = \begin{cases} \dfrac{\left(\sum PW + ID\right)\times\left(-1\right)^{SS}}{\sum PW + \left|\sum NW\right| + 1} & \sum PW > \left|\sum NW\right| \\ 0 & \sum PW = \left|\sum NW\right|. \\ \dfrac{\left(\sum NW - ID\right)\times\left(-1\right)^{SS}}{\sum PW + \left|\sum NW\right| + 1} & \sum PW < \left|\sum NW\right| \end{cases} \quad \textbf{(5)}$$

In formula (5), *PW* represents the weight of positive sentiment words, *NW* represents the weight of negative sentiment words, *ID* represents the increase or decrease of sentiment intensity, and *SS* represents the number of sentiment shifter or coordinating conjunction "but". Among them, *ID* includes sentiment comparison, dynamic modality, increase and decrease word, *ID* is 0.5 if the sentiment intensity is strengthened, *ID* is -0.5 if the sentiment strength is weakened.

(3) Judgment of Sentiment Tendency of English Essays

We can calculate the sentiment score of English essays according to the sentiment score of sentences. The calculation formula is shown in (6):

$$Essay\ score = \begin{cases} \dfrac{\sum PSS}{\sum PSS + \left|\sum NSS\right| + 1} & \sum PSS > \left|\sum NSS\right| \\ 0 & \sum PSS = \left|\sum NSS\right|. \\ \dfrac{\sum NSS}{\sum PSS + \left|\sum NSS\right| + 1} & \sum PSS < \left|\sum NSS\right| \end{cases} \quad \textbf{(6)}$$

In formula (6), *PSS* represents positive sentence scores, *NSS* represents negative sentence scores. Then we divided English essays into strong positive, positive, weakly positive, weakly negative, negative, strong negative, which is the sentiment tendency of English essays, according to the sentiment scores of English essays. The pseudo-code for the calculation of sentiment score of English essays is shown in Algorithm 2:

Algorithm 2. Pseudo-code Of Sentiment Score Calculation for English Essays

```
INPUT: English essays
OUTPUT: Sentiment score of English essays
 BEGIN
     For (i=0; i<sentences. Length)
       For (j=0; j<words. Length)
          If (word='JJ' or word='NN' or word='VB' or word='RB') Then
             Word SentiValue=calculate SentiValue (word);
          If (words have "but" or sentiment shifter) Then
             Sentence SentiValue=-sentence SentiValue;
          If (words have increment degree word) Then
             Sentence SentiValue++;
         If (words have decrement degree word) Then
            Sentence SentiValue--;
        ELSE
           Essay SentiValue = 0;
    Essay SentiValue = calculate SentiValue (sentence SentiValue);
    return essay SentiValue;
  END
```

4 Experiment

4.1 Dataset

The experiment data in this paper comes from CLEC, the learner's English corpus, which is the first publicly available learner's English corpus. We select English essays from CLEC's CET4 sub-dataset and TEM sub-dataset as experiment data, and select 5000 English essays from each sub-dataset, a total of 10000 English essays for training. As for the test set, in the opinion analysis test, we select 500 English essays that on-topic and 500 English essays that off-topic from each sub-dataset. A total of 2000 English essays as a test set. In the sentiment analysis test, we select 1000 English essays that on-topic in opinion analysis test as a test set. In the thought expression analysis test, we select 2000 English essays in the opinion analysis test as a test set. The opinion analysis of English essays is mainly to judge whether the English essay is off topic. The opinion analysis test dataset includes two parts of model essays dataset and essays to be analyzed dataset. The model essays are on-topic English essays which have been manually evaluated. Model essays are used as reference object in experiment. The essays that to be analyzed refers to the English essays that has not been evaluated, and it needs to be compared with the model essays to analyze and evaluate whether the English essays is on-topic or not.

We choose another five model essays as model essays dataset, greatly reduce the number of model essays required compared with other methods 30 model essays.

Table 1 shows the sentiment statistics of the datasets CET4 and TEM. As can be seen from Table 1, the proportion of neutral sentiment in the dataset TEM is higher than CET4, while the proportion of positive and negative sentiment in the dataset TEM is lower than CET4. It is probably because the English essays in the TEM are more subtle.

Table 1. Sentiment Statistics for two data sets

Data	Positive	Neutral	Negative
CET4	82.1%	3.2%	14.7%
TEM	76.2%	12.3%	11.5%

4.2 Preprocess and Parameter Set

We perform some pre-processing on two datasets, (1) delete non-English letters, punctuation marks, (2) segmentation of English essays, (3) delete stop words and perform stem, It should be noted that sentiment shifter often plays a key role and cannot be removed, (4) we mark POS (parts of speech) for words using the Stanford POS Tagger.

In SP-LDA method, α =1, β =1 in initial state. The number of themes is 5 and the number of output opinion words is 5. We calculate the approximate value of parameters in SP-LDA using the Gibbs sampling algorithm. The maximum number of Gibbs sampling times is 10000.

4.3 Result Analysis and Discussion

4.3.1 Experiment Analysis of Opinion Analysis

Comparison with baselines: SP-LDA proposed is compared with LDA, SLDA and LDA with bigram in this paper using standard evaluation indicators, accuracy, precision, recall, and F1. We test 1000 English essays on-topic and 1000 English essays off-topic in opinion analysis respectively. The results of LDA, SLDA, LDA with bigram, and SP-LDA are compared as follows under the condition that 5 English model essays. Table 2 and Table 3 are the experiment results on the test set CET4 and TEM, respectively.

Table 2. Comparison of results on CET4 test set

Method	On-topic				Off-topic			
	A	P	R	F1	A	P	R	F1
LDA [3]	41.71%	41.68%	41.73%	41.70%	48.21%	48.33%	48.25%	48.29%
SLDA [5]	73.69%	75.36%	74.38%	74.86%	79.21%	81.14%	79.98%	80.55%
LDA+bigram [11]	74.22%	75.71%	75.77%	75.74%	79.97%	80.90%	80.74%	80.82%
SP-LDA (Our)	74.66%	76.42%	75.36%	75.88%	80.28%	82.17%	81.07%	81.61%

Table 3. Comparison of results on TEM test set

Method	On-topic				Off-topic			
	A	P	R	F1	A	P	R	F1
LDA [3]	41.77%	41.74%	41.71%	41.72%	46.27%	46.31%	46.29%	46.30%
SLDA [5]	73.98%	74.68%	76.21%	75.44%	77.95%	79.28%	78.57%	78.92%
LDA+bigram [11]	74.65%	74.82%	77.03%	75.91%	78.69%	79.07%	79.57%	79.32%
SP-LDA (Our)	75.02%	75.46%	77.08%	76.26%	79.65%	80.29%	79.63%	79.96%

Table 2 and Table 3 show that SP-LDA performs better than LDA, SLDA and LDA with bigram in both the on-topic and off-topic in test sets of CET4 and TEM. The off-topic has higher accuracy and F1 than on-topic, which may be because essays off-topic contain less information of model essays, it is relatively easier to distinguish. Comparing the results of CET4 and TEM, the accuracy, precision, recall and F1 of CET4 and TEM are similar under four methods in the on-topic, while the F1 of TEM is lower 1.99%, 1.63%, 1.50% and 1.65% than CET4 under four methods in the off-topic, respectively, the reason may be that the diversity and implication of the describe in the TEM essays, it is relatively difficult to detect.

The influence of the number of model essays: The opinion analysis test requires a certain number of English model essays as reference. We take 5, 10, 15, 20, 25 and 30 English model essays respectively to test the influence of the number of model essays on the experiment results, and the results are shown in Fig. 3:

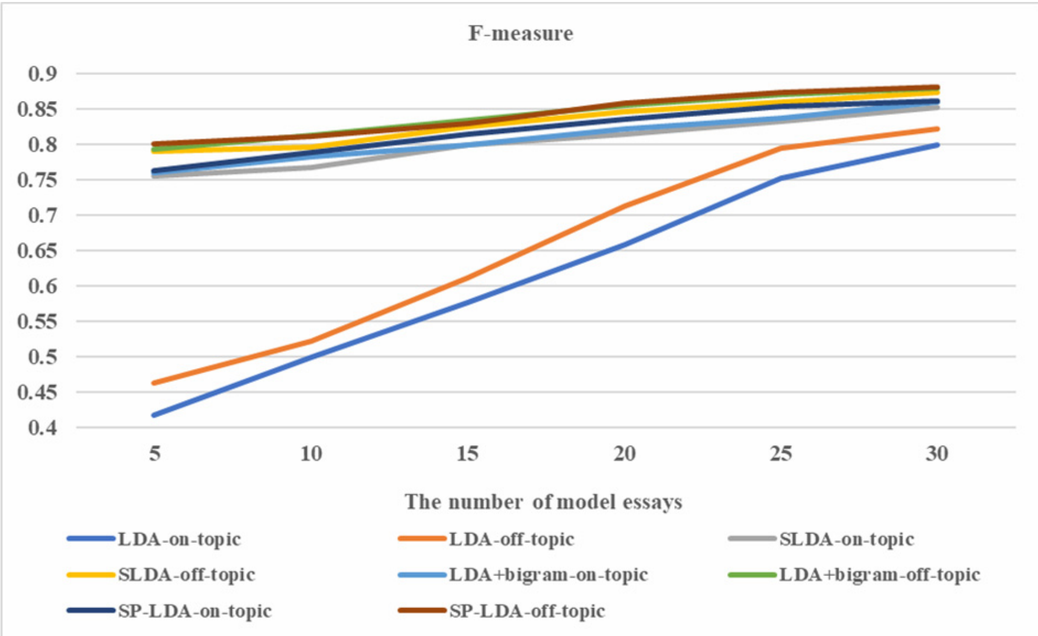

Fig. 3. The influence of the number of model essays on the experiment results

Fig. 3 shows that F1 of SP-LDA in this paper is higher than LDA, SLDA, and LDA with bigram in both the on-topic and off-topic. In addition, the F1 of LDA, SLDA, LDA with bigram and SP-LDA has different degrees of improvement as the number of model essays increases. SP-LDA has reached a higher F1 value when the number of model essays is 5, which is improved on the basis of the SLDA method, and the result is almost close to the F1 of LDA when the number of model essays is 25. Therefore, it is reasonable to select 5 English model essays as references.

4.3.2 Experiment Analysis of Sentiment Analysis

The influence of different sentiment lexicons: The current common English sentiment lexicons include MPQA, Bing Liu Opinion Lexicon, Inquirer, LIWC, NRC Hashtag Sentiment Lexicon, NRC Sentiment140 Lexicon, SentiWordNet3.0, SenticNet4, SenticNet5 etc. We choose the most popular sentiment Lexicons recently for sentiment analysis and comparison: SentiWordNet3.0, SenticNet4, SenticNet5. The sentiment score of words is directly given in SenticNet4 and SenticNet5, but it is necessary to calculate the comprehensive sentiment score of words in sentiwordnet3.0 because each word may have multiple positive and negative scores due to different POS and semantics. Its calculation formula is as shown in (7):

$$word\ score=\frac{\sum_{i=1}^{n}\frac{Positive\ score\ \text{-}\ Negative\ score}{i}}{\sum_{i=1}^{n}\frac{1}{i}}\ . \tag{7}$$

In formula (7), n is the number of synonyms sets of sentiment word, i is the serial number in the synonym set of sentiment word. The positive score of sentiment words is the positive score of sentiment words in the synonym set of sentiwordnet3.0. The negative score of sentiment words is the negative score of sentiment words in the synonym set of sentiwordnet3.0. The sentiment analysis results in different lexicons shown in Table 4:

Table 4. Sentiment Analysis Results in Different Lexicons

Lexicon	Positive				Negative			
	A	P	R	F1	A	P	R	F1
SentiWordNe3	59.53%	72.24%	58.83%	64.84%	58.56%	58.56%	68.82%	63.27%
SenticNet4	73.72%	80.27%	78.22%	79.23%	71.75%	77.06%	78.99%	78.01%
SenticNet5	67.21%	69.62%	67.11%	68.34%	71.35%	72.24%	74.77%	73.48%

Table 4 shows that the sentiment analysis results in different Lexicons. In both positive and negative aspects, SenticNet5 has higher accuracy and F1 than SentiWordNet3.0 and SenticNet4. SenticNet5 contains more sentiment words category and sentiment phrases, which results in better accuracy and F1.

The influence of different Parts of Speech: We test the influence of extracting adjectives and adverbs as sentiment words, sentiment shifter (contains coordinate conjunction "but", SS), increase and decrease word (contains sentiment comparison and dynamic modality, ID) on sentiment analysis of English essays in 1000 on-topic essays with SenticNet5 and the results are shown in Table 5:

Table 5. The Influence of Parts of Speech on Sentiment Analysis

Variation	Positive				Negative			
	A	P	R	F1	A	P	R	F1
Trad-Method	59.53%	72.24%	58.83%	64.84%	58.56%	58.56%	68.82%	63.27%
Only-adj/adv	73.72%	80.27%	78.22%	79.23%	71.75%	77.06%	78.99%	78.01%
Only -SS	67.21%	69.62%	67.11%	68.34%	71.35%	72.24%	74.77%	73.48%
Only -ID	70.29%	71.62%	69.32%	70.45%	68.26%	68.39%	70.98%	69.66%
ALL	81.92%	89.98%	84.01%	86.89%	78.32%	81.12%	87.95%	84.39%

Table 5 shows that the sentiment analysis method proposed in this paper has a great improvement in the positive and negative aspects compared with the traditional method. The traditional method does not consider sentiment shifter, coordinate conjunction, sentiment comparison, dynamic modality, increase, and decrease word. Relatively speaking, the accuracy of negative sentiment analysis is slightly lower. The reason may be that the author usually adopts euphemistic implicit words, or by the way that expresses positive and negative sentiment at the same time and then denies positive sentiment to express the negative sentiment. Table 5 also shows that the extraction of adjectives and adverbs as sentiment words has the greatest influence on our method, which illustrate that adjectives and adverbs can represent the author's subjective feelings. Sentiment shifter has a greater impact on the judgment of negative sentiment than positive sentiment, which is because the author directly uses positive sentiment words to express positive sentiment, while negative sentiment is usually expressed by denying positive sentiment.

Comparison with baselines: We compared the sentiment analysis method proposed in this paper with the two baseline methods:

Asghar, M. Z. [20] et al. considered emoticons, modifiers, and domain specific terms to analysis sentiment of reviews, we call this method EMD for comparison purposes; Saif, H. [21] et al. proposed a Lexicon-based method——SentiCircles, which considered the co-occurrence patterns of words in different contexts. The experiment comparison results are shown in Table 6:

Table 6. Comparison of Different Sentiment Analysis Methods

Methods	Positive				Negative			
	A	P	R	F1	A	P	R	F1
EMD [20]	81.74%	84.21%	89.15%	86.61%	78.03%	82.57%	80.00%	81.26%
SentiCircles [21]	82.11%	90.16%	83.71%	86.82%	77.15%	80.72%	83.17%	81.93%
Our method	81.92%	89.98%	84.01%	86.89%	78.32%	81.12%	87.95%	84.39%

As can be seen from the experiment data in Table 6, comparing with EMD and SentiCircles, our method performs better both in positive and negative categories. This is because we consider more influencing factors on sentiment analysis of English essays. SentiCircles has higher accuracy in the positive category, as SentiCircles considered the co-occurrence patterns of words in different contexts to identify sentiment words pairs more accurately. While EMD has higher accuracy than SentiCircles in the negative category, it shows a relatively better performance in English essays because of considering modifiers and domain specific terms.

The influence of different themes: There are five themes in sentiment analysis test set: technology, environment, work, education, and the Internet. We analyze and compare these themes to obtain the author's sentiment on them.

Fig. 4 shows the sentiment analysis and comparison on the five hot topics of technology, environment, work, education, and Internet. We can know that people express more positive sentiment about the environment and education in these five topics. 72.5% of the essays express positive sentiment on technology, while 22.3% think that science and technology bring convenience to people and cause lots of trouble at the same time. People also express much negative sentiment about the network, and think that the network brings more negative effects to people. People express more worries and concerns about the hot topic of work.

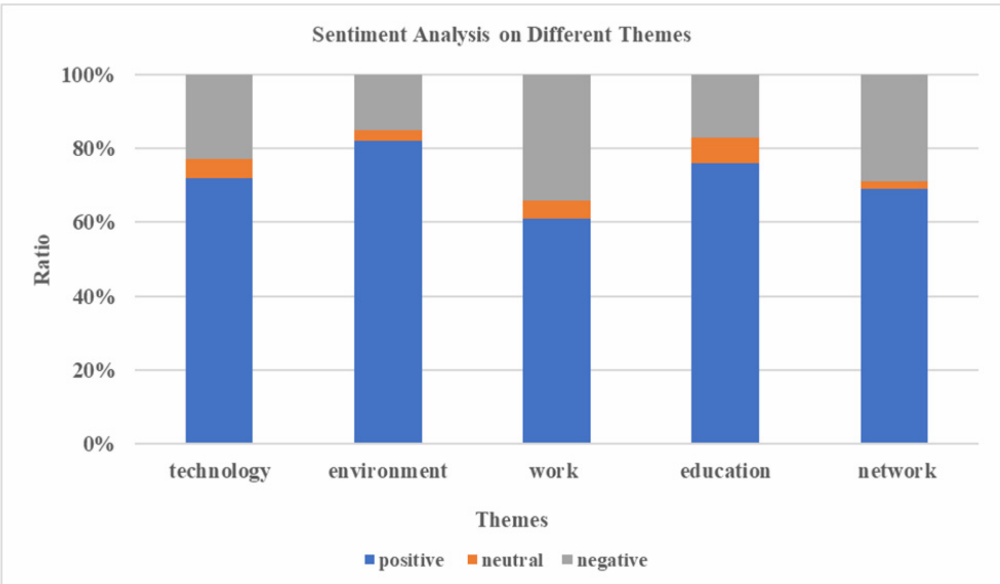

Fig. 4. Sentiment analysis and comparison of different themes

4.3.3 Experiment Analysis of Thought Expression

The full score of thought expression in English Essays is 20, including two parts of the relevant degree of opinion and sentiment tendency. The experiment results of the two parts have been analyzed separately. This section will analyze the thought expression through experiment We compare the manual score and automatic score to analyze the performance of the proposed method in this paper. The weight of opinion analysis and sentiment analysis is 1 / 2 in thought expression analysis respectively. The experiment comparison results are shown in Fig. 5:

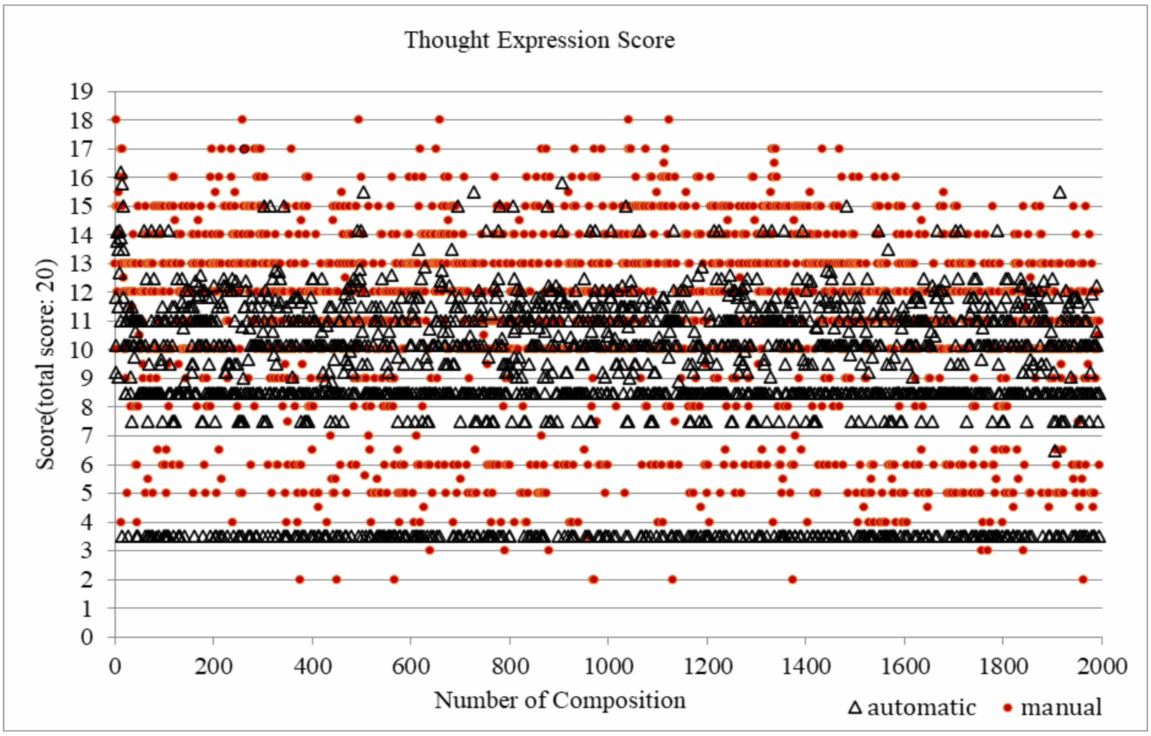

Fig. 5. Experimental analysis of thought expression

Fig. 5 shows that the scores of automatic corrections are denser than manual, and the scores of manual corrections are generally higher than automatic correction. The reason may be that people are easily affected by perceptual thinking. An essay will be given a score according to the actual situation even if the expression of it is unclear, while there are no such factors in the automatic correction. We can gain the average score of automatic corrections is 9.11, and manual corrections is 11.07, the difference between them is 1.96, which is an acceptable range. Although there is still a gap between our results and manual correction, it is a relatively great result because the process of analyzing thought expression of English essays is very subjective and only 5 English model essays as references in opinion analysis test.

5 Conclusion

In this paper, we propose to combine sentiment and opinion to solve the problem that the automatic English essays correction system lacks analysis of thought expression of English essays. Our method can automatically analyze the relevant degree of opinion and sentiment tendency of English essays, and then analyze the thought expression of English essays by combing sentiment and opinion to obtain thought expression scores and comment. The results show that the method we proposed has a good performance in analyzing thought expression of English essays by experiment analysis.

In the future, we intend to judge the relevance degree of opinion in English essays by using the Knowledge Graph method to further reduce the number of English model essays required and conduct fine-grained sentiment analysis on English essays.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62066009, No. 61662012).

References

[1] M. Liu, Design of Intelligent English Writing Self-evaluation Auxiliary System, Informatica 43(2)(2019).

[2] J. Zhou, Application investigation and technological development of intelligent product in college English teaching, Journal of Physics: Conference Series. IOP Publishing 1237(2)(2019).

[3] Blei, M. David, Y. Andrew, I. Michael, Latent dirichlet allocation, Journal of Machine Learning Research (2003) 993-1022.

[4] H. Jelodar, Y. Wang, C. Yuan, Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey, Multimedia Tools and Applications 78(11)(2019)15169-15211.

[5] G. Huang, Z. Sun, C. Fan, J. Liu, A Sentence Level Corrected LDA Topic Model for Chinese English Learners, in: Proc. ITITS 2018 Information Technology and Intelligent Transportation Systems (2019) 206-214.

[6] K. Ravi, V. Ravi, A survey on opinion mining and sentiment analysis: tasks, approaches and applications, Knowledge-Based Systems 89(2015) 14-46.

[7] J. Qiang, Z. Qian, Y. Li, Y. Yuan, X. Wu, Short Text Topic Modeling Techniques, Applications, and Performance: A Survey, arXiv: Information Retrieval (2019).

[8] G. Xu, X. Wu, H. Yao, F. Li, Z. Yu, Research on Topic Recognition of Network Sensitive Information Based on SW-LDA Model, IEEE Access 7(2019) 21527-21538.

[9] K. Bastani, H. Namavari, J. Shaffer, Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints, Expert Systems with Applications 127(2019) 256-271.

[10] Y. Gao, X. Wei, X. Zhang, W. Zhuang, A Combinational LDA-Based Topic Model for User Interest Inference of Energy Efficient IPTV Service in Smart Building, IEEE Access 6(2018) 48921-48933.

[11] M. Helal, M. Mouhoub, Topic Modelling in Bangla Language: An LDA Approach to Optimize Topics and News Classification, Computer and Information Science 11(4)(2018).

[12] K. Porter, Analyzing the DarkNetMarkets subreddit for evolutions of tools and trends using LDA topic modeling, Digital Investigation 26(2018) S87-S97.

[13] C. Huang, Q. Wang, D. Yang, Topic mining of tourist attractions based on a seasonal context aware LDA model, Intelligent Data Analysis 22(2)(2018) 383-405.

[14] H. Nabli, R. B. Djemaa, I. A. B. Amor, Efficient cloud service discovery approach based on LDA topic modeling, Journal of Systems and Software 146(2018) 233-248.

[15] C. SG Khoo, S. B. Johnkhan, Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons, Journal of Information Science 44(4)(2018) 491-511.

[16] S. Trinh, L. Nguyen, M. Vo, Lexicon-based sentiment analysis of Facebook comments in Vietnamese language, Springer (2016) 263-276.

[17] J. Thompson, B. Leung, M. Blair, M. Taboada, Sentiment analysis of player chat messaging in the video game StarCraft 2: Extending a lexicon-based model, Knowledge-Based Systems 137(2017) 149-162.

[18] F. H. Khan, U. Qamar, S. Bashir, A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet, Knowledge and Information Systems 51(3)(2017) 851-872.

[19] C. Dhaoui, C. Webster, L. Tan, Social media sentiment analysis: lexicon versus machine learning, Journal of Consumer Marketing 34(6)(2017) 480-488.

[20] M. Asghar, A. Khan, S. Ahmad, Lexicon-enhanced sentiment analysis framework using rule-based classification scheme, PloS one 12(2)(2017).

[21] H. Saif, Y. He, M. Fernandez, H. Alani, Contextual semantics for sentiment analysis of Twitter, Information Processing and Management 52(1)(2016) 5-19.

[22] Y. Takano, Y. Iijima, K. Kobayashi, Improving Document Similarity Calculation Using Cosine-Similarity Graphs, Springer (2019) 512-522.

[23] B. Liu, Sentiment analysis: Mining opinions, sentiments, and emotions, Cambridge University Press, 2015.