

Research Article

Off-Topic Detection of Business English Essay Based on Deep Learning Model

Yiting Zhu 

Wuhan University of Bioengineering, Wuhan 430415, China

Correspondence should be addressed to Yiting Zhu; janezhu@whsw.edu.cn

Received 18 September 2021; Revised 6 October 2021; Accepted 9 October 2021; Published 30 October 2021

Academic Editor: Chin-Ling Chen

Copyright © 2021 Yiting Zhu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The automatic scoring system of business English essay has been widely used in the field of education, and it is indispensable for the task of off-topic detection of essay. Most of the traditional off-topic detection methods convert text into vector representation of vector space and then calculate the similarity between the text and the correct text to get the off-topic result. However, those methods only focus on the structure of the text, but ignore the semantic association. In addition, the traditional detection method has a low off-topic detection effect for essays with high divergence. In view of the above problems, this paper proposes an off-topic detection method for business English essay based on the deep learning model. Firstly, the word2vec model is used to represent words in sentences as word vectors. And, LDA is used to extract the vector of topic and text, respectively. Then, word vector and topic word vector are spliced together as the input of the convolutional neural network (CNN). CNN is used to extract and screen the features of sentences and perform similarity calculation. When the similarity is less than the threshold, the paper also maps the topic and the subject words in the coupling space and calculates their relevance. Finally, unsupervised off-topic detection is realized by the clustering method. The experimental results show that the off-topic detection method based on the deep learning model can improve the detection accuracy of both the essays with low divergence and the essays with high divergence to a certain extent, especially the essays with high divergence.

1. Introduction

Automatic composition scoring is a process of autonomously participating in composition scoring using computer-related technologies [1], such as natural language processing and machine learning. Before scoring the composition language logic, word composition, and semantic level, testing the degree of suitability between the composition and the topic is the first step in the scoring system. For example, the topic of the test question is “English composition writing.” Although the composition to be tested uses beautiful words, logical language, and novel perspectives, the topic is quite different from “English composition writing.” Then, the score for such a composition is naturally not high according to the actual situation. Therefore, off-topic detection of composition is the first “checkpoint” of automatic composition scoring, which conforms to the rules and conventions of manual scoring. Off-topic detection is of

great significance to the accuracy, practicality, and fairness of automatic composition scoring.

Business English composition off-topic detection refers to the use of computer-related technology to automatically detect whether the subject content of the composition deviates from the topic requirements. The core of off-topic detection is to calculate the similarity between texts. For example, the similarity between the sample essay and the essay to be tested can be used to detect whether the essay is off-topic. As an auxiliary technology for automatic scoring of composition, researchers at home and abroad have also carried out a number of studies. With the development of natural language processing technology, related research will receive more attention. Literature [2] applies off-topic detection methods in other literature studies to Portuguese detection and compares them in an experiment on a public corpus of 2164 papers related to 111 prompts. Literature [3] proposed an unsupervised detection method for off-topic

papers based on goals and reference hints, which uses the semantic difference between the similarity of the article and the target prompt and the similarity with the reference prompt to calculate the topic score. This can better distinguish topic articles and off-topic articles. This method realizes an unsupervised paper off-topic detection system without a large amount of training data. Literature [4] proposed a method to verify user response using the off-topic paper detection method used in automatic paper evaluation. In the proposed C-BGRU Siamese architecture, the convolutional layer learns from the embedding vector of the word and captures the contextual features of the word n-gram. The bidirectional gated recurrent unit (BGRU) is used to access previous and subsequent context representations. Literature [5] proposes two methods to better compare the semantic similarity between the article to be tested and the article. One method is to use WordNet, a hand-compiled thesaurus, to measure text similarity. The other method uses real-valued vectors learned from a large corpus to represent words in the text before calculating the distance. Literature [6] proposed a method based on the convolutional neural network (CNN), which can make better use of the semantic information contained in the report text to speed up the retrieval process. The proposed method uses the graph embedding method to enhance the word representation by capturing the semantic relationship information from the medical ontology. And, the CNN model is adjusted to calculate the similarity between the report pairs to determine the target report-paired with overlapping body parts. Experiments show that this method can realize the semantic similarity detection of medical texts. Literature [7] expresses each short text as two dense vectors: the former uses word-to-word similarity construction based on pretrained word vectors, and the latter uses word-to-word similarity construction based on external knowledge sources. They also developed a preprocessing algorithm that links related named entities together and performs word segmentation to preserve the meaning of phrasal verbs and idioms. Experimental data show that the interdependent representation of short-text pairs is effective and efficient for semantic text similarity tasks. Literature [8] uses the Biterm-LDA model to extract the subject words of the title and the article and combines it with Doc2vec to check the combined subject and semantics. Secondly, the author proposes a threshold calculation method based on the center of the topic composition under different topic composition. Finally, the ROC curve is used to find the optimal threshold for each type of topic composition and then the off-topic composition is judged based on the optimal threshold. However, for actual examinations, it is more and more common for test questions to require authors to use divergent thinking to create questions. Divergent thinking creation is also a major focus and trend in the field of educational examinations in the future. However, due to the higher divergence, there are two problems in the following two problems: the type of sample essay cannot be fully covered, and the result of the similarity test with the title is poor.

In response to the abovementioned problems, this paper proposes a method of off-topic detection of business English composition based on the deep learning model. This method

first preprocesses the article. Word2vec is used to train the word vector, and LDA is used to extract the subject word vector of the topic and the text, respectively. The two vectors are combined and used as the input of CNN for feature extraction. Then, the distributed vector representation model is used to calculate the similarity between them. When the similarity is greater than a certain threshold, it is considered that the subject word of this content meets the topic, and the subsequent correlation calculation is not carried out. The correlation of other subject words is calculated and represented by vectors. Finally, the clustering method is used to complete the detection. Experiments show that this method can improve the detection effect of compositions with high topic divergence.

2. The Related Work

2.1. Convolutional Neural Network. Convolutional neural network (CNN) is a feedforward neural network with convolution calculation. As one of the representative algorithms of deep learning, CNN has been used in various fields. CNN imitates the visual perception mechanism of living organisms, so it can be used for supervised learning and unsupervised learning. As a deep neural network, CNN can combine some filters with the input and can guarantee the translation invariance to deal with the size changes in the input data. Each filter in CNN is a weight vector that can be trained. The feedforward network is interchanged between the convolutional layer and the maximum pooling layer. The top layer consists of many sparse or fully connected layers. Finally, there is the final decision-making layer or classification layer. Deep CNN training is usually done through supervised learning and is faster than training other types of neural network models. The small amount of weight in CNN makes it more effective than other neural network-based feature extraction methods.

The output of each neuron in CNN is calculated according to its input and the weight and deviation of neurons in previous layers of the network structure. The following equation is used to update the weight and deviation of each layer, respectively.

$$\begin{aligned}\Delta w_i(s+1) &= -arw_i - \beta \frac{\partial C}{\partial w_i} + d \Delta w_i(s) \\ \Delta b_i(s+1) &= -a \times \frac{\partial C}{\partial b_i} + d \Delta b_i(s),\end{aligned}\tag{1}$$

where s represents the update step. $a = \theta/n$. θ represents the regularization parameter. n indicates the training sample. r stands for learning rate. w_i and b_i are the weight and deviation of neuron i , respectively. d is the total number of momenta. C is the cost function.

2.2. LDA Keyword Extraction. Latent Dirichlet allocation (LDA) was proposed by Blei, Ng, and Jordan in 2003. Since the model is simple and effective, LDA is currently applied in the field of text mining, including text topic recognition, text classification, and text similarity calculation. LDA is a master

model that mines potential topics from a given document [9]. The emergence of LDA is to solve the problem that TF-IDF can only measure document similarity from word frequency. The two documents are similar situations because few or no words exist in both documents together. LDA is often used for semantic mining to identify potential topic information in documents. In the topic model, a topic represents a concept, an aspect. LDA is represented as a series of related words and is the conditional probability of those words.

LDA is an unsupervised generation model. The premise of the model is that the generation of each word in each article is “selecting a certain topic with a certain probability and selecting a certain word from the topic with a certain probability” [10]. So, the probability of each word appearing in each document is calculated as follows:

$$P(\text{documentation}|\text{words}) = \sum_{\text{topic}} P(\text{words}|\text{topic})P(\text{topic}|\text{document}). \quad (2)$$

The LDA topic model can also be represented as a probability graph. The representation of LDA is shown in Figure 1.

2.3. Distributed Vector Representation. Word vector is originally represented by the one-hot representation method. However, this traditional approach can cause serious dimensional disaster and generate word vectors that fail to represent their meanings. In view of the above shortcomings, distributed vector representation method emerged. In the case of large vector dimensions, each word can be represented by the distributed weight of the element. Each dimension of the vector represents an eigenvector, acting on all terms, rather than simply a one-to-one mapping between elements and values. This method abstractly represents the “meaning” of a word. At present, the open source tool word2vec implemented by Google using the model proposed by Mikolov et al. in 2013 is widely used in the distributed representation of word vectors [11, 12]. It can train word vector quickly and effectively [13].

Word2vec includes two important models: Continuous Bag-of-words (CBOW) model and Skip-Gram model. Both models include three layers: input layer, hidden layer, and output layer. The input layer of the CROW model is the bag-of-words vector representation of the context words of the target words. The output layer is the bag-of-words vector representation of the target words. The CROW model predicts the semantics of target words through context and is suitable for small databases. The input layer of the skip-gram model is the word bag vector representation of the target word. The output layer is the bag-of-words vector representation of the context words of the target words. Therefore, the skip-gram model is suitable for data sets with a large amount of data.

3. Methods

3.1. Feature Representation. The feature representation of essays is a vectorization of short text data to reflect the features of texts with high representativeness and high

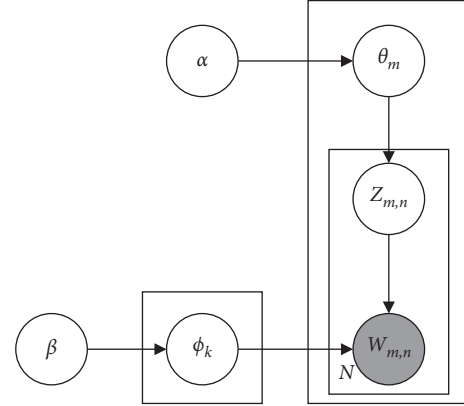


FIGURE 1: The structure of CNN. m represents the article. α represents the parameter of Dirichlet distribution of the prior distribution of the topic distribution of each article. β represents the parameter of Dirichlet distribution prior distribution of word distribution for each topic. θ_m represents the topic distribution of article m . ϕ_k represents the word distribution for the m th topic. N represents the total number of words in the article. $Z_{m,n}$ represents the topic to which the n th word in the m th article belongs. $W_{m,n}$ stands for the n th word in the m th article.

computational value. From the level of word granularity, word2vec mines the meaning of words to express the fine semantics of the text. From the level of text granularity, LDA builds the topic distribution of the text through a probability model, focusing on the overall semantic expression of the text. Through vector splicing, they build a feature matrix containing word meaning and semantics to ensure the integrity of composition features from two levels.

Word2vec can quickly construct the word vector form of words. The value of each dimension of word vector represents a feature with certain semantic and grammatical interpretation. Skip-gram uses the way of jump combination to learn word rules, which can better meet the needs of sparse features of short texts:

$$P(F_{t-k}F_{t-k+1}F_{t+k-1}F_{t+k}|F_t), \quad (3)$$

where P is the probability of the occurrence of F_t constructed by jump combination. F_t is any term that appears in the set of articles. And, $F_{t-k}F_{t-k+1}F_{t+k-1}F_{t+k}$ is the jump combination of F_t . During word vector training, N -dimensional vector representation of any word is output.

$$E = \begin{bmatrix} e_{11} & e_{12} & e_{13} & \cdots & e_{1N} \\ e_{21} & e_{22} & e_{23} & \cdots & e_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ e_{k1} & w_{k2} & e_{k3} & \cdots & e_{kN} \end{bmatrix}, \quad (4)$$

where E is any article in the article set and e_{kN} is the weight eigenvalue of the text term k .

LDA topic model analyses the implied semantic topic of the text through the topic model of prior probability [14]. The semantic topic reflects the implied semantics of the text and is a direct extraction of the deep features of the text. LDA is used to train short-text corpus and output text-topic

matrix and topic-word matrix. The word vector and the topic vector are superimposed by vector splicing to form a new input matrix. This matrix contains both the semantic features of the word and the overall semantic features.

$$E_{\text{new}} = E \oplus Z, \quad (5)$$

where \oplus is the vector splicing operation. The input matrix E represents the word vector corresponding to the text. Z represents the topic distribution vector corresponding to the text. This matrix is used as input data of CNN convolution layer. Convolution layer and pooling layer can be used for feature extraction and classification, so as to obtain the result of text similarity.

3.2. Article Similarity Calculation of Depth Model

3.2.1. Convolution Layer. The purpose of convolution layer is to extract higher-order features implied by arbitrary input matrix E_{k+1N} by convolution operation. The convolution sequences of different convolution kernels can be obtained through sliding calculation of convolution kernels E_{k+1N} at different heights on $H^{l,a}$. The convolution sequence constitutes the feature surface $h^{l,a}$, where l represents the height of the convolution kernel. a represents the text vector dimension. $h^{l,a}$ is actually the result set of the inner convolution of the input $H^{l,a}$ with E_{k+1N} and the bias. Considering that the preprocessed short-text data contain fewer quantitative feature words, and the convolution window is set as 5. And, the convolution step is 1.

$$h^{l,a} = f(x * H^{l,a} + b^{l,a}), \quad (6)$$

where f is the activation function tanh, which is used to smooth the convolution result. Its purpose is to introduce nonlinearity into neural network and ensure the curve relationship between input and output. The result of the convolution layer is the set of feature surfaces through multiple convolution kernels.

3.2.2. Pooling Layer. Pooling layer is to downsample the set of high-dimensional feature surfaces to prevent overfitting and improve computing performance. The pooling process typically divides the input features into subregions of several sizes. Each subregion is pooled. The common pooling method is max-pooling, and the pooled values are displayed.

The feature surface extracted by each convolution kernel is operated. Finally, each convolution kernel corresponds to a value. When these values are spliced together, a new feature quantity representing the sentence is obtained.

3.2.3. Output Layer. The output layer mainly realizes the connection between the convolution layer and the pooling layer. After many cycles of convolution layer and pooling layer, several dynamic values are finally realized to represent the similarity representation of modified sentences.

For the two documents D_m and D_n , the document D_m is divided into $B_m^1, B_m^2, \dots, B_m^p$ p sentences, and the document D_n is divided into $B_n^1, B_n^2, \dots, B_n^q$ q sentences. $\text{Sim}(D_m, D_n)$

represents the percentage of contents in a document D_m that are similar to D_n . $\text{Sim}(D_n, D_m)$ represents the percentage of document D_n content similar to document D_m . In this study, $\text{Sim}(D_m, D_n)$ and $\text{Sim}(D_n, D_m)$ are not necessarily equal. For example, $D_m = \text{"Hello, it's a nice day today,"}$, $D_n = \text{"Hi, it's a nice day, want to go out and play?"}$. Here, the whole content of D_m is exactly the same as part of D_n . In this study, $\text{Sim}(D_m, D_n) = 1$. However, only part of the D_n content is the same as that in D_m , so $\text{Sim}(D_n, D_m) < 1$.

Suppose document D_m contains s sentences $B_m^1, B_m^2, \dots, B_m^s$. If the sentence similarity between D_m and document D_n exceeds a certain threshold, then $\text{Sim}(D_m, D_n)$ in this study is calculated according to equation (8):

$$\text{Sim}(D_m, D_n) = \frac{\sum_{i=1}^s \text{length}(D_m^i)}{\text{length}(D_n)}. \quad (7)$$

Suppose the document D_n contains t sentences $B_n^1, B_n^2, \dots, B_n^t$. If the sentence similarity with document D_m exceeds a certain threshold, $\text{Sim}(D_n, D_m)$ in this study is calculated according to

$$\text{Sim}(D_n, D_m) = \frac{\sum_{i=1}^s \text{length}(D_n^i)}{\text{length}(D_m)}. \quad (8)$$

3.3. Coupled Spatial Model (CSM). At present, most of the research on off-topic detection of composition is based on the calculation of text similarity, so as to judge whether it is off-topic. However, for a composition with a high divergence, only using text similarity to judge whether it is off topic is too strict to the tangency. As a result, many compositions that meet the requirements of the topic are judged to be off-topic compositions. This will affect the accuracy of off-topic detection.

Therefore, the concept of text relevance is put forward in this paper. Text relatedness is defined as the correlation between texts calculated by using the spatial distance generated by the coupling space vector established by co-occurrence correlation, conditional correlation, and coupling correlation between texts [15]. Text relevance refers to the fact that words in the same context or in the same article are often related, even though they are not necessarily similar in meaning. Here are the definitions.

Definition 1 (co-occurrence correlation). Co-occurrence correlation represents the probability of word pairs appearing in the same article at the same time. n represents the number of articles in the article set $A = \{a_1, a_2, \dots, a_n\}$. m represents the number of words contained in the article set $S = \{s_1, s_2, \dots, s_m\}$. In this paper, Tanimoto coefficient is used to calculate the co-occurrence frequency between pairs of subject words.

$$R_{co}(s_i, s_j) = \frac{1}{|A'|} \sum_{x \in A} \frac{T_{xi} T_{xj}}{\|T_{xi}\|^2 + \|T_{xj}\|^2 - T_{xi} T_{xj}}, \quad (9)$$

where T_{xi} and T_{xj} , respectively, represent the TF-IDF weight of a group of words to s_i and s_j in the article. $|A'|$ represents

the number of articles whose TF-IDF weights for s_i and s_j are not 0. The co-occurrence relatedness of subject words to s_i and s_j is the sum of co-occurrence relatedness of word pairs to all articles in the article set.

Definition 2 (conditional correlation). When one item w_i of the word pair appears in article E_x , the probability of another item w_j appearing in the document is expressed as

$$R_{\text{condition}}(s_i, s_j) = \begin{cases} 1, \\ \frac{R_{\text{co}}(s_i, s_j)}{\sum_{i=1, i \neq j}^{m-1} R_{\text{co}}(s_i, s_j)}, \end{cases} \quad (10)$$

where m represents the number of words contained in the set. R_{co} represents the degree of co-occurrence of words s_i and s_j .

Definition 3 (coupling correlation). When there is at least one word s_k in article set A , so that $R_{\text{condition}}(s_i, s_k)$ and $R_{\text{condition}}(s_k, s_j)$. Then, words have a coupling correlation with s_i and s_j . The higher the number of related words between s_i and s_j , the stronger the coupling correlation between word pairs is. So, the coupling relevance of words to s_i and s_j can be expressed as

$$R_{\text{coupling}}(s_i, s_j) = \begin{cases} 0, & i = j, \\ \frac{1}{|L|} \sum_{\forall w_k \in L} R_{\text{coupling}}(s_i, s_j | s_k), & i \neq j, \end{cases} \quad (11)$$

where L indicates the number of words related to s_i and s_j . When there is no such correlation between word pairs s_i and s_j , the coupling correlation is 0.

In order to express the correlation between word pairs more reasonably, the conditional correlation and coupling correlation are linearly weighted. The greater the value of $R(s_i, s_j)$, the greater the correlation between s_i and s_j . By training the corpus and considering the inclusion of complete semantics, we can get the universal semantic matrix $E_{e_{ij}}$ containing all word pairs.

$$R(s_i, s_j) = \begin{cases} 1, & i = j, \\ \alpha \cdot R_{\text{condition}}(s_i, s_j) + (1 - \alpha) \cdot R_{\text{coupling}}(s_i, s_j) & i \neq j. \end{cases} \quad (12)$$

It takes a long time to calculate the relatedness of words in coupled space compared with the traditional calculation method. Therefore, the traditional similarity calculation equation can be directly used to calculate the word pairs with high semantic similarity. The rest of the words enter the coupling space for correlation calculation. This will improve the detection efficiency of the algorithm to a certain extent. Therefore, it is also very important to select the parameter of word similarity ψ , which is involved in the method proposed in this paper. Different similarity condition ψ will affect the accuracy of experimental results.

Therefore, condition ψ with different similarity was selected as the comparative experiment, and the optimal solution was selected as the similarity condition parameter of the experiment in this paper. Here, the influence of similarity condition ψ on accuracy is more obvious. Articles with higher divergence were selected for the experiment, and the specific experimental results are shown in Figure 2.

As can be seen from the above experimental results, the accuracy rate increases with the increase of ψ . The difference in accuracy before $\psi = 0.85-0.97$ is small and tends to plateau. Considering the running time, this paper selects $\psi = 0.9$ as the semantic similarity condition for the experiment.

3.4. Algorithm Description. Given the set of article topics $G = \{g_1, g_2 \dots g_v\}$ and the anthology $O = \{o_{11}, o_{12} \dots o_{1n_1}; o_{21}, o_{22} \dots o_{2n_2}; \dots; o_{v1}, o_{v2} \dots o_{vn_2}\}$, where v represents the number of topics contained in the article set and o_{ij} represents the j th article under title g_i . The specific algorithm process is as follows.

The performance of off-topic detection is improved by calculating the correlation between word pairs in coupled space. The architecture of off-topic detection model of business English composition based on deep learning model is shown in Figure 3.

4. Analysis of Experimental Results

4.1. Experimental Data and Experimental Parameter Setting. In order to verify the effectiveness of the proposed method, this paper selects three datasets from Kaggle's composition scoring competition. There are two main sources of digression essays in the data set. An off-topic composition selected from the original data set that has been marked by humans as a low score. The other one is randomly selected from other topics. The detailed description of the whole experiment is shown in Table 1. The experimental data shown in this article is obtained by averaging the experimental data of 3 data sets.

All experiments in this paper are carried out using Python 3.5.2 on 64-bit Windows10 operating system with Intel (R) core (TM) i7-4510 CPU @2.0 GHz and 8GB RAM. CNN parameter settings are as follows. The convolution kernel is 150. The convolution kernel window is 5. The discard rate is 0.5, and the learning rate is $1e-3$. The initialization parameter ratio of LDA is shown in Table 2.

4.2. The Evaluation Index. Suppose the article under a certain topic is denoted as $w_{i1}, w_{i2}, \dots, w_{in_j}$. The divergence value of the topic is measured by the similarity between the articles.

$$\text{div} = \frac{1}{\text{Group}} \sum_{\substack{1 \leq p \leq n_j \\ 1 \leq q \leq n_j}} \text{Sim}(w_{ip}, w_{iq}), \quad (13)$$

where Group represents the number of pairwise combinations of w_{ip} and w_{iq} . $\text{Sim}(w_{ip}, w_{iq})$ represents the similarity

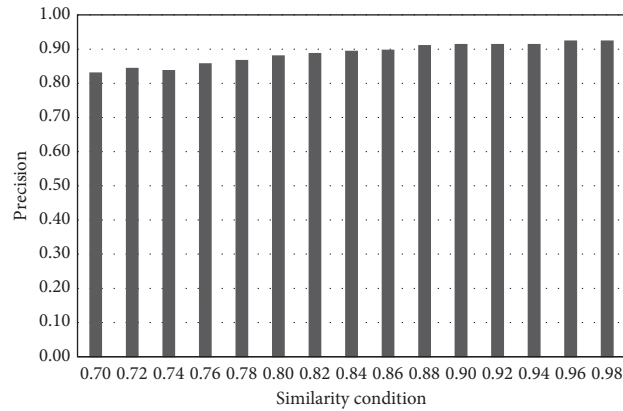


FIGURE 2: The relationship between accuracy and similarity condition setting.

Input: topic set G and article set O . The K value of the class cluster is set to 2
 Output: segmented clusters of articles.

- (a) Preprocessing the original text, including word segmentation, removing punctuation, removing blank characters, removing stop words, etc.
- (b) Enter the short text and the corresponding article respectively. Word vector is obtained by using the trained word2vec model.
- (c) LDA topic model is used to obtain the subject words of the topic and the article respectively. The topic selects the topic probability top-5 topic words. The article selects the topic probability top-15 topic words. A subject-word matrix is obtained by combining the subject word and word vector.
- (d) As the input of CNN, the topic-word matrix will be used to calculate the similarity between test articles and article set vectors.
- (e) When the similarity is greater than ψ , the essay to be tested is considered to meet the requirements of the question. No correlation calculation is required.
- (f) When the relative similarity is greater than ψ , the coupling spatial model is used to calculate the correlation. In step (e), all the relevant degrees of keywords selected were set to 1, and the pan-semantic matrix $E_{o_{ij}}$ was obtained. The rows of the matrix represent the article subject words. Columns represent subject headings.
- (g) Take the maximum vector value of each column as an element in the vector representation of this article, and get a new vector.
- (h) Cycle the above 7 steps until all articles to be tested are represented as distributed vectors.
- (i) Input article vector and output article cluster by k -means clustering method.

ALGORITHM 1: Off-topic detection algorithm for business English composition based on deep learning.

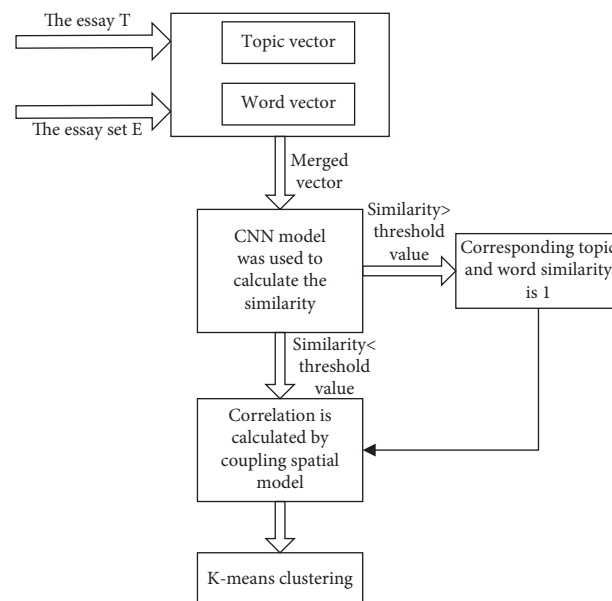


FIGURE 3: The off-topic detection architecture based on deep learning model.

TABLE 1: Description of our datasets.

Data	Essay target prompt	On-topic essays	Off-topic essays
1	Facing business challenges at holiday inn worldwide	1726	328
2	What plan you have for future development?	1753	301
3	What is important when shopping online?	1756	276

TABLE 2: The initialization parameter ratio of LDA.

α	β	Niters Gibbs	K
50/ K	0.01	1000	100

between w_{ip} and w_{iq} , which is calculated using the TF-IDF method. The higher the divergence value div , and the lower the divergence of the article title.

The evaluation indicators used in this article are the retrieval accuracy P (precision), the recall rate R (recall), and the F1 measurement values commonly used in information detection.

4.3. Analysis of Experimental Data. The proposed method is compared with the existing methods in the values of P , R , and $F1$. The off-topic detection method of LDA model in literature [15]. Unsupervised off-topic detection model based on mixed semantic space (MSS) in literature [16]. The method based on text clustering (TC) in literature [17]. First, the performance of the four methods on precision P is compared, as shown in Figure 4.

The abscissa in Figure 4 shows the divergence of the topic. Divergence indicates the topic range of the topic. The greater the divergence is, the more difficult the off-topic detection will be. The ordinate represents the retrieval accuracy. The experimental results show that, with the increase of the divergence of the topic, the four methods show a downward trend on the whole. The method based on text clustering decreased rapidly. The precision of the proposed method is basically the same as the other three methods in the region with low divergence. However, the precision of this paper is the highest in the region with high divergence.

In Figure 5, the recall rate performance of the four methods under different divergence of topics is illustrated. It can be seen that when the divergence of the topic is 0.2, and the recall rate of the text clustering method reaches 0.94. The recall rate of mixed semantic space is 0.85. The recall rate of LDA model is 0.83. The recall rate of the method based on divergence and dynamic threshold is 0.77. When the divergence of the problem is low, text clustering method performs best. But on the whole, the method presented in this paper is the most stable and performs well when divergence is high.

As shown in Figure 6, the value of $F1$ shows a downward trend with the increase of divergence. The detection method based on the deep learning model proposed in this paper does not show obvious $F1$ value in the region with low divergence. However, this method has the advantage in the high divergence area. The reason why the method in this paper has better performance in composition detection with higher divergence is because the method in this paper uses LDA and

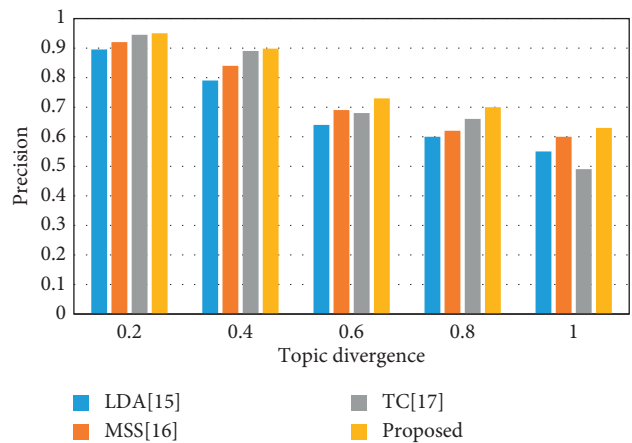


FIGURE 4: The relationship between divergence and precision.

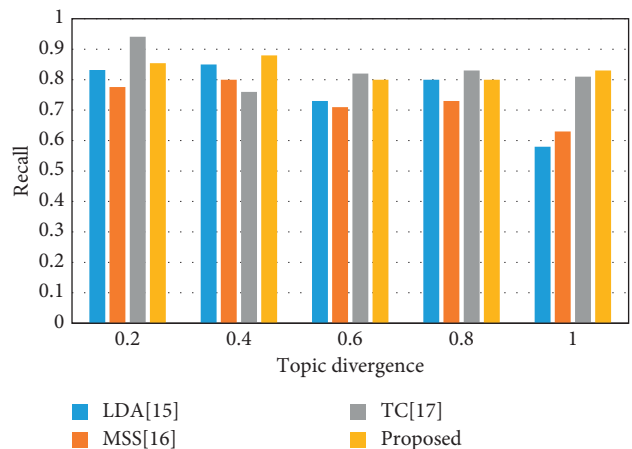


FIGURE 5: The relationship between divergence and recall.

word2vec for training to construct a feature matrix containing word meaning and semantics to ensure the integrity of the composition characteristics from two levels. The combined features are used as the input of CNN for similarity calculation. In addition to the similarity comparison, this paper also uses the coupling space for correlation detection, further improving the accuracy of off-topic detection.

By comparing the above three groups of experimental results, the method proposed in this paper is more stable than other methods in terms of accuracy, recall rate, and

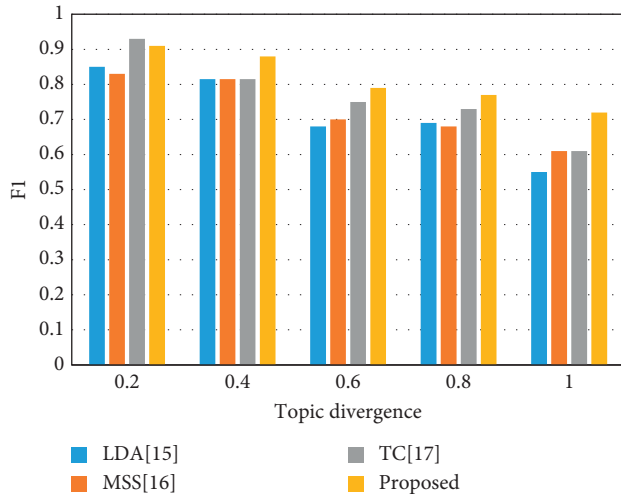


FIGURE 6: The relationship between divergence and F1.

more objective $F1$ value. It is shown that this method is not sensitive to the index of divergence of the topic and can still perform well. In other words, this method can show relatively stable and accurate judgment in detecting whether the article is off-topic when facing the requirements of the article with high divergence in practical application. This will be more in line with the trend of more divergent and creative thinking in actual exams. In summary, the performance of this method on real data sets is not obvious in areas with low divergence of the title. However, compared with traditional methods, it has also improved, and all indicators perform better in areas where the divergence of the question is higher.

5. Conclusion

An off-topic detection method for business English composition based on the deep learning model is proposed in this paper. The innovation of this method lies in the combination of word vector at word granularity level and topic vector at the semantic granularity level. As the input of the deep CNN model, feature extraction and similarity calculation are carried out. Then, the feature mapping value is added to the coupling space to increase the “understanding” of the detection system, so as to improve the index of the off-topic detection method. In particular, they performed better in areas with higher divergence. However, the method in this paper is greatly limited by the data set of training coupled space. How to reduce the influence of data set on the detection method will be one of the main directions of future research. In addition, further research on reducing the calculation time of the algorithm is needed.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

Acknowledgments

This work was supported by the Research on Innovation of Application-Oriented Talents Cultivating Mode for International Trade Major Led by Enterprises under the Condition of “Work-integrated Study” (project no. 2017GB128).

References

- [1] J. Qian and Z. Wang, “A comparative study on the effectiveness of online automatic scoring of English writing in juku correction network and mosoteach APP under blended teaching mode,” *Journal of Physics: Conference Series*, vol. 1533, no. 4, Article ID 042082, 2020.
- [2] G. Passero, R. Ferreira, and R. L. S. Dazzi, “Off-topic essay detection: a comparative study on the Portuguese language,” *Revista Brasileira de Informática na Educação*, vol. 27, no. 3, pp. 177–190, 2019.
- [3] X. Li, Q. Wen, and K. Pan, “Unsupervised off-topic essay detection based on target and reference prompts,” in *Proceedings of the 2017 13th International Conference on Computational Intelligence and Security (CIS)*, pp. 465–468, IEEE, Hong Kong, China, 2017.
- [4] T. M. Tashu, “Off-topic essay detection using C-BGRU siamese,” in *Proceedings of the 2020 IEEE 14th International Conference on Semantic Computing (ICSC)*, pp. 221–225, IEEE, San Diego, CA, USA, 2020.
- [5] L. Liu, “A comparative study of different text similarity measures for identification of off-topic student essays,” *Boletim Técnico/Technical Bulletin*, vol. 55, no. 11, pp. 602–606, 2017.
- [6] T. Zheng, Y. Gao, F. Wang et al., “Detection of medical text semantic similarity based on convolutional neural network,” *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, pp. 1–11, 2019.
- [7] H. T. Nguyen, P. H. Duong, and E. Cambria, “Learning short-text semantic similarity with word embeddings and external knowledge sources,” *Knowledge-Based Systems*, vol. 182, Article ID 104842, 2019.
- [8] P. Liu, J. Liu, X. Ma, and J. Zhou, “Off-topic detection model based on biterm-LDA and doc2vec,” in *Proceedings of the 2019 3rd High Performance Computing and Cluster Technologies Conference*, pp. 157–161, ACM, Guangzhou, China, 2019.
- [9] H. Jelodar, Y. Wang, C. Yuan et al., “Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey,” *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 15169–15211, 2019.
- [10] S. W. Kim and J. M. Gil, “Research paper classification systems based on TF-IDF and LDA schemes,” *Human-Centric Computing and Information Sciences*, vol. 9, no. 1, pp. 1–21, 2019.
- [11] D. Jatnika, M. A. Bijaksana, and A. A. Suryani, “Word2Vec model analysis for semantic similarities in English words,” *Procedia Computer Science*, vol. 157, pp. 160–167, 2019.
- [12] M. Grohe, “Word2vec, node2vec, graph2vec, x2vec: towards a theory of vector embeddings of structured data,” in *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, pp. 1–16, ACM, Portland, OR, USA, 2020.
- [13] A. Khatua, A. Khatua, and E. Cambria, “A tale of two epidemics: contextual Word2Vec for classifying twitter streams during outbreaks,” *Information Processing & Management*, vol. 56, no. 1, pp. 247–257, 2019.
- [14] C. Yan, X. Chang, M. Luo et al., “Self-weighted robust LDA for multiclass classification with edge classes,” *ACM Transactions*

- on Intelligent Systems and Technology (TIST)*, vol. 12, no. 1, pp. 1–19, 2020.
- [15] Q. Qu, Y. Zhao, and R. Cui, “Computer-aided off-topic composition detection,” in *Proceedings of the 2nd International Conference on Culture, Education and Economic Development of Modern Society (ICCESE 2018)*, pp. 155–158, Atlantis Press, Moscow, Russia, 2018.
- [16] G. Huang, J. Liu, C. Fan, and T. Pan, “Off-topic English essay detection model based on hybrid semantic space for automated English essay scoring system,” *MATEC Web of Conferences*, vol. 232, p. 01035, 2018.
- [17] J. Wang, “Application of text clustering in automatic scoring of college,” in *Proceedings of the 2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, pp. 598–603, IEEE, Guangzhou, China, 2020.