Classification of Learning Sentiments of College Students Based on Topic Discussion Texts of Online Learning Platforms

https://doi.org/10.3991/ijet.v17i24.35951

Mingzhe Niu(✉)
College of Innovation and Entrepreneurship, Zibo Vocational College, Zibo, China
11017@zbvc.edu.cn

Abstract—Depicting the online learning process of student users from multiple angles can help implement deep learning and effectively improve their online learning quality, and it's a practical and very meaningful work to mine the data burying in the topic discussion texts of online learning platforms so that useful information could be extracted and attained to help teachers better understand students' learning sentiments and assist students to know of the learning status of their peers. However, in existing conventional sentiment analysis methods, the sample data with uncategorized tags are still labelled manually, and such work is usually time consuming and inefficient. In view of these defects, this paper aims to study the classification of college students' learning sentiments based on the topic discussion texts of online learning platforms. In the beginning, this paper gave the overall structure of the proposed college student Learning Sentiment Classification (LSC) algorithm, and discussed the similarity between the topic discussion content and the teaching content. Then, this paper proposed to integrate Convolution Neural Network (CNN) with the Long-Short Term Memory (LSTM) network to build the said LSC model, so as to merge the advantages of the two and improve the accuracy of learning sentiment rating. After that, embedding layers of static words and non-static words were introduced into the proposed model for the purpose of realizing the mining of specific textual information while enhancing the semantic expression ability of the words. At last, experimental results verified the effectiveness of the proposed model.

Keywords—online learning, topic discussion, text mining, learning sentiment analysis

1 Introduction

On online learning platforms, topic discussion is a common and effective interaction method of remote education that can strengthen teacher-student connections, student-student mutual influence, and the integration of online teaching and online learning [1–8]. To implement deep learning and improve the quality of online learning, it is a necessary work to assess students' academic performance, objective achievement, and knowledge and skill level, so as to depict their online learning process from multiple angles, and the topic discussion on the online learning platforms has made this work possible [9–15].

As online learning platforms have become increasingly popular worldwide, topic discussion is now an indispensable part of students' online learning process. Topic discussion is a good means for students to express their views and listening to the ideas of others, with the progress of learning stages, the data volume of topic discussion texts accumulates and grows constantly [16–20], and various types of interactive information co-exist in the topic discussion forums of online learning platforms, containing both the deep level interactions and the simple comments or information exchange. Therefore, it's a practical and very meaningful work to mine the data burying in the topic discussion texts so that useful information could be extracted and attained to help teachers better understand students' learning sentiments and assist students to know of the learning status of their peers.

Jiao et al. [21] discussed the argumentation characteristics of socio-scientific issues in online video learning websites, they employed the content analysis method to conduct coding and discourse analysis on 112 argumentation discourses in the discussion forums of online video learning websites, and figured out the argumentation characteristics of high-likes comments and the different argumentation performances of high-quality and low-quality arguments. Patni et al. [22] argued that the content and strategy online discussion plays an important role in blended learning, but the contents and strategies currently implemented have not been able to contribute optimally, so the authors proposed this paper as a developmental research with design-based research approach. The paper set goals and roles for online discussion, prepared content that facilitates students to learn independently, motivated them to have the confidence to hold discussions, and let teachers involve in the discussions, then the rewards and punishments were applied based on the given scores and extra tasks to enhance positive competitiveness, and the assessments were carried out via online quizzes. To study the effects of grouping strategies on asynchronous online discussion, Li et al. [23] designed a quasi-experimental research which took 178 graduate students as subjects to compare the differences in student participation and social interactions between whole-class and small-group discussions, and among groups composed of self-selected acquaintances, partial acquaintances, and randomly-assigned strangers, and the statistical results attained from learning analytics and social network analysis support the superiority of small discussion groups self-formed by acquaintances in increasing learning participation and social interactions. Barman et al. [24] qualitatively investigated the interactions between learners, as well as between learners and teachers in MOOCs. In their paper, the community of inquiry was taken as the analytical framework to unveil how the discussions made in the online environment may connect to the course participants' learning processes, and the preliminary findings suggest that the interactions taking place in the discussion forums primarily concern issues regarding course structure.

Existing studies on the sentiment analysis methods mostly talk about building sentiment word dictionaries or the quantification of sentiments, these methods tend to cover as much sentiment words as possible or figure out the tendency of the sentiments hidden in the target texts, however, the sample data with uncategorized tags are still labelled manually, and such work is usually time consuming and inefficient. In view of these defects, this paper studied the classification of college students' learning sentiments based on the topic discussion texts of online learning platforms. In the second chapter, this paper gave the overall structure of the proposed LSC algorithm and discussed the similarity between the topic discussion content and the teaching content. In the third chapter, this paper proposed to integrate CNN with LSTM network to build the said LSC model, so as to merge the advantages of the two and improve the accuracy of learning sentiment rating. After that, embedding layers of static words and non-static words were introduced into the proposed model for the purpose of realizing the mining of specific textual information while enhancing the semantic expression ability of the words. At last, experimental results verified the effectiveness of the proposed model.

2 Calculation of the semantic similarity of topic discussion texts

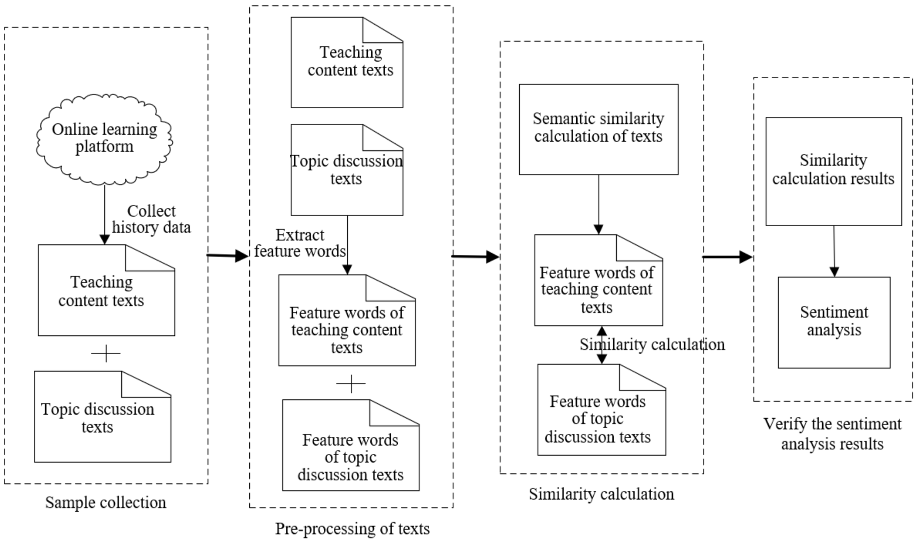

Fig. 1. Overall structure of the LSC algorithm

To accurately classify the learning sentiments of college students, this paper proposed the LSC algorithm based on the mined data of topic discussion texts of online learning platforms, and Figure 1 gives the overall structure of the algorithm, specific steps include collect samples of topic discussion texts, pre-process the topic discussion texts, calculate the similarity of teaching content texts, and verify the sentiment analysis results, etc.

In order to mine the data of the topic discussion texts of online learning platforms based on semantics, at first, the similarity between the discussion text content and the teachers' teaching content should be discussed.

The research objective of this paper is to score the learning sentiments of students for a certain online course during their online learning process, so this paper selected the textual sematic similarity between the teaching content feature word set and the topic discussion text content feature word set of the online learning platforms as the reference for rating the learning sentiments of students during their online learning, that is, the higher the similarity of students' topic discussion texts for a certain online course, the more positive their for participating in the learning of this online course.

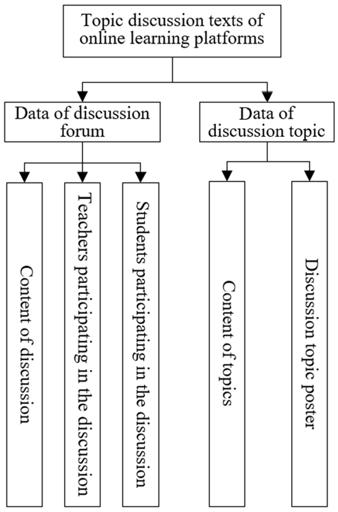

Fig. 2. Data structure of topic discussion texts

According to the requirements of the experiment, this paper took each single discussion topic as a unit to process the texts, and the data of the discussion topic and discussion forum were separated. Figure 2 gives the data structure of the topic discussion texts. For the discussion topics, the data were attained from 2 parts: the topic poster, and the topic content; for the discussion forum, the data were attained from 3 parts: students participating in the discussion, teachers participating the discussion, and the discussion content.

Therefore, in this paper, the score of students' learning sentiments was converted into the comprehensive score of the similarity between teaching content texts and topic discussion texts. The meaning of the comprehensive score is the weighted sum of the similarity values of the discussion texts of each topic. Assuming: there're m discussion topics about an online course; q_i represents the weight value corresponding to the sematic similarity result $S\text{-}Sim_i$ of the i-th discussion topic, then the following formula calculates the comprehensive score ZF_X:

$$ZF_X = \sum_{i=1}^{m} S - Sim_i \times q_i \tag{1}$$

Figure 3 shows the principle for calculating the word similarity. As can be seen from the figure, the calculation of word similarity is the basis for calculating the sematic similarity between the teaching content text feature word set and the topic discussion text feature word set, and its calculation accuracy is very important for the rating the learning sentiments of students.

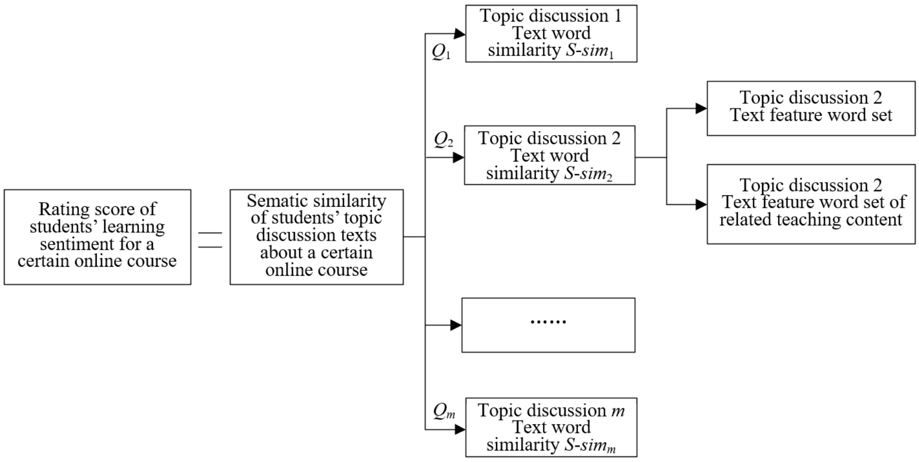

Fig. 3. Calculation principle of word similarity

Assuming: $\gamma_i (1 \leq i \leq 4)$ represents an adjustable parameter; $S\text{-}Sim_1(R_1, R_2)$ represents the similarity of the first basic sememes between two real word concepts R_1 and R_2; $S\text{-}Sim_2(R_1, R_2)$ represents the similarity of other basic sememes between R_1 and R_2; $S\text{-}Sim_3(R_1, R_2)$ represents the similarity of relationship sememes between R_1 and R_2; $S\text{-}Sim_4(R_1, R_2)$ represents the similarity of relationship symbols between R_1 and R_2; then the following formula calculates the similarity between these two real word concepts:

$$S - Sim(R_1, R_2) = \sum_{i=1}^{4} \gamma_i S - Sim(R_1, R_2) \quad (2)$$

The adjustable parameter γ satisfies the following requirements:

$$\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = 1, \gamma_1 \geq \gamma_2 \geq \gamma_3 \geq \gamma_4 \quad (3)$$

The following calculates the similarity between teaching content texts and topic discussion texts. Assuming: a student has participated in the discussion of o topics about the online course; then for this student X, the comprehensive similarity of the topic discussion texts of this online course could be calculated using this formula:

$$S - Sim_X = \sum_{l=1}^{o} S - Sim_{Xl} \alpha_l \quad (4)$$

Assuming: $S\text{-}Sim_{Xl}$ represents the similarity of discussion texts of the l-th topic participated by student X; α_l represents the weight of the k-th discussion topic about this online course, then α_l needs to satisfy the following constraint:

$$\sum\nolimits_{l=1}^{o} \alpha_l = 1 \quad (5)$$

Next, *S-Sim*$_{Xl}$, the similarity of discussion texts of the l-th topic participated by student X was calculated; assuming: there're n unrepeated feature words in discussion text TE_a of student X under a certain topic, denoted as $p_1, p_2, \ldots p_n$; TE_b represents the course knowledge point text corresponding to this topic, and there're m different feature words in TE_b, denoted as $p_1, p_2, \ldots p_m$; *S-Sim*$(p_1)_{max}$ represents the maximum similarity of p_1, then the similarity between texts TE_a and TE_b can be calculated by the following formula:

$$S-Sim_{Xl} = S-Sim_{(a,b)} = \frac{\sum_{i=1}^{n} S-Sim(p_i)_{max} \times q_{ai} \times q_{bj}}{n} \tag{6}$$

From *S-Sim*$(p_{a1}p_{b1})_{max}$, *S-Sim*$(p_{a1}p_{b2})_{max}$, ..., *S-Sim*$(p_{a1}p_{b1})_{max}$, the maximum value was selected and taken as the maximum similarity *S-Sim*$(p_1)_{max}$ of the feature word t_1p_1. Figure 4 gives the strategy for selecting the maximum similarity of feature words.

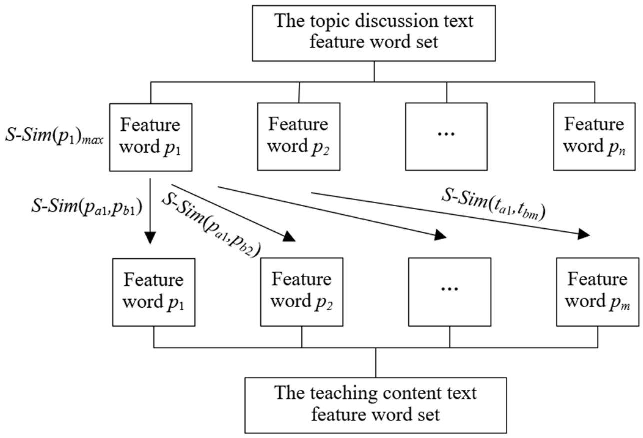

Fig. 4. Strategy for selecting the maximum similarity of feature words

One thing should be noted is that the feature words of the topic discussion texts of online learning platforms should be of multiple uses; assuming: g_i represents the frequency of p_i in TE_a, then the following formula calculates the weight Q_{ai} of p_i in TE_a:

$$q_{ai} = \frac{g_i}{\sum_{i=1}^{m} j_i} \tag{7}$$

Similar to Q_{ai}, Q_{bj} also represents the weight of p_j in TE_b. p_j is the corresponding feature word in TE_b when p_i reaches the maximum similarity, then the weight Q_{bj} of p_j in TE_b is given by the following formula:

$$q_{bj} = \frac{g_j}{\sum_{i=1}^{m} j_i} \tag{8}$$

3 Classification of learning sentiments based on topic discussion

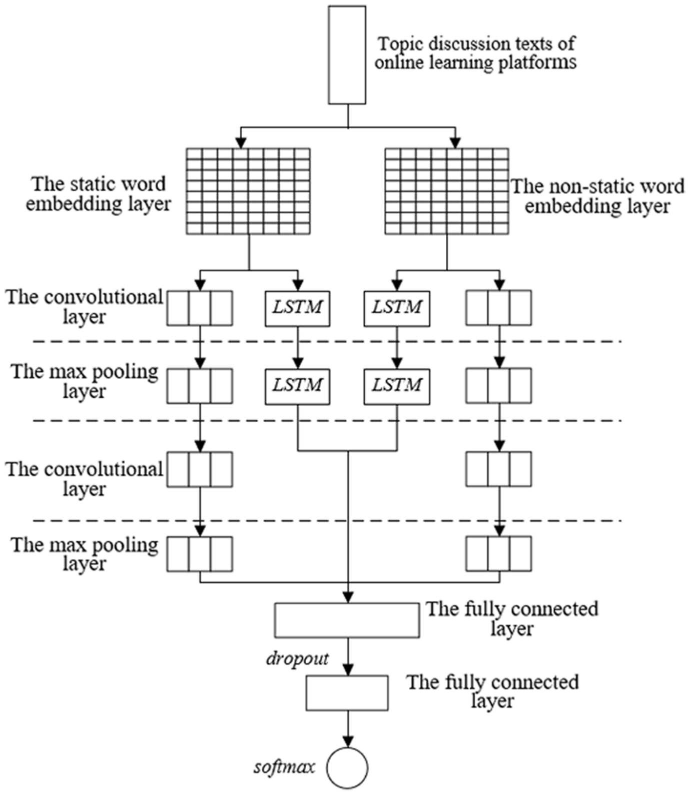

Fig. 5. Overall structure of the combined model

In learning sentiment analysis, strong feature words in a topic discussion text of online learning platforms often determine the sentiment tendency of the entire topic discussion text. CNN is not suitable for dealing with long distance dependencies, and the LSTM network with input gate and forget gate set for information update is suitable for processing sequence information data, so this paper integrated the two to merge their advantages and improve the accuracy of the rating of learning sentiments. In addition, existing word-embedding models generally have defects in their textual expression ability, so this paper introduced the static and non-static word embedding layers into the constructed model, thereby realizing the mining of specific information in the texts while enhancing the sematic expression ability of the words. Figure 5 shows the overall structure of the combined model.

In this paper, the convolution operation of the CNN module was realized by the sliding of the one-dimensional filter on the topic discussion text word vector matrix; assuming: $a_{i:i+f-1}$ represents the vector from the i-th word to the $i+f-1$-th word in the sentence; Q represents the filter; f represents its height; m represents the width which is the dimensionality of the word vector; κ represents the bias term; g represents the *Relu* function; $\otimes$ represents the point-by-point multiplication, then there is:

$$r_i = g(q \otimes a_{i:i+f-1} + \kappa) \tag{9}$$

After the convolution operation shown in the above formula, the feature sequence R of the topic discussion texts of online learning platforms could be attained, that is, $R = [r_1, r_2, ..., r_{n-f+1}]$. In order to extract the features of topic discussion texts with different granularities, different sizes were assigned to the filter height f, then, through the maximum pooling processing, the representative features of the topic discussion texts of online learning platforms were attained:

$$\hat{t}_k = max[r_{2i-1}, r_{2i}] \tag{10}$$

Next, the second round of convolution and pooling operations were performed to further extract the features of the topic discussion texts of online learning platforms so as to get a wider receptive field. The output of the first operation of the pooling layer was taken as the input, and the calculation formula of the second round of convolution is:

$$r_i' = g(q \otimes \hat{t}_{i:i+f-1} + \kappa) \tag{11}$$

In the calculation result of the above formula, the feature sequence R' contains l features, namely $R' = [r_1', r_2', \ldots, r_l']$, the representative feature is the maximum value selected by the second pooling layer from R', then there is:

$$z = max[r_1', r_2', ..., r_l'] \tag{12}$$

The feature vectors of topic discussion texts extracted by the CNN based on the static and non-static embedding layers were respectively represented by z_1, z_2, z_3 and z_4, z_5, z_6; and the feature vectors of topic discussion texts extracted by the LSTM network based on the static and non-static embedding layers were respectively represented by k_1 and k_2.

Next, before z_1, z_2, z_3, z_4, z_5, z_6 and k_1, k_s were delivered to the fully connected layer, the following splicing operation was performed:

$$N = SO(z_1, z_2, z_3, z_4, z_5, z_6, k_1, k_2) \tag{13}$$

Assuming: Q_m represents the weight matrix; κ_n represents the bias; g represents the *Relu* activation function, then the fully connected layer integrated the feature vector N that had completed the splicing:

$$E = g(Q_m N + \kappa_n) \tag{14}$$

Assuming: Q_e represents the weight matrix; κ_y represents the bias; then through the *Softmax* function, the classification result of the rating of students' learning sentiment was attained from E as:

$$t = Softmax(Q_e E + \kappa_e) \tag{15}$$

Assuming: c_i represents the output value of the i-th node of the output layer; Z represents the number of output layer nodes that characterizes the rating level of learning sentiments, the following formula gives the expression of the *Softmax* function:

$$t_i = \frac{o^{c_i}}{\sum_{z=1}^{z} o^{c_z}} \tag{16}$$

In topic discussion texts with sentiment tendency, there're certain differences in students' learning sentiments carried by texts of different time series. This paper introduced the self-attention mechanism into the constructed combined model to focus on local key points of the topic discussion texts.

The input of the self-attention layer is $F' = [f_1', f_2', \ldots, f_n']$, which is the output of the LSTM network module, and the self-attention can be described as a task-related query w. The output of the self-attention layer is the mapping of a series of (l-u) pairs. The matrices constituted by the query vectors, the key vectors, and the value vectors are respectively represented by $W = [w_1, w_2, \ldots, w_m]$, $L = [l_1, l_2, \ldots, l_m]$, and $U = [u_1, u_2, \ldots, u_m]$, then the three matrices were subject to spatial linear mapping to attain vectors w_i, l_i, and u_i. Assuming Q^W, Q^L, and Q^U represent the weight matrices of the linear mapping, then the calculation formulas of the matrices are:

$$W = Q^W F' \tag{17}$$

$$L = Q^L F' \tag{18}$$

$$U = Q^U F' \tag{19}$$

To determine the similarity between $W = [w_1, w_2, \ldots, w_m]$ and $L = [l_1, l_2, \ldots, l_m]$, this paper defined a scoring function as follows which is required to have the name dimensionality of w and l:

$$SF(w,l) = w^T l \tag{20}$$

To cope with the problem that the dot product operation would increase with the increase of data dimensionality, this paper set a scaling factor θ, then there is:

$$SF(w,l) = \frac{w^T l}{\theta} \tag{21}$$

When the fractions of multiple w_i and l_i vectors need to be determined, all the vectors can be replaced by corresponding matrices W and L. The output of the self-attention layer is given by the following formula:

$$SA(W,L,U) = U * soft\max(SF(W,L)) \tag{22}$$

The output of the self-attention layer is the result output by the static and non-static word embedding layers after processed by the LSTM network, the attained feature vectors are respectively represented by x_1 and x_2; then x_1 and x_2 were spliced with the feature vectors extracted by the CNN, then there is:

$$N = SO(z_1, z_2, z_3, z_4, z_5, z_6, x_1, x_2) \tag{23}$$

N was input into the fully connected layer, and the final classification result of students' learning sentiments was output by the output layer.

4 Experimental results and analysis

To verify the effectiveness of the proposed sematic similarity calculation method of topic discussion texts, this paper employed actual examples for the verification. The topic discussion text a is "Advanced mathematics is a compulsory mathematics course for students majoring in science and engineering disciplines but not majoring in mathematics, the main contents include: series of numbers, limits, calculus, analytic geometry of space, and linear algebra". The teacher's teaching content is "Higher mathematics is a basic subject formed by calculus, algebra, geometry, and the interdisciplinary content between them, it is a basic exam subject for students majoring in engineering, science, and finance". Then, the proposed similarity calculation method was applied to calculate the sematic similarity between the teaching content text feature word set and the topic discussion text feature word set of the online learning platforms, and the results are listed in the Table 1.

Table 1. Calculation results of semantic similarity of feature word sets

	Advanced Mathematics	**Calculus**	**Algebra**	**Geometry**
Advanced mathematics	1.257	0.015	0.074	0.017
Science and engineering	0.016	0.025	0.041	0.036
Required course	0.032	0.041	0.069	0.014
Math class	0.018	0.069	0.034	0.069
Series of numbers	0.021	0.035	0.958	0.052
Limits	0.016	0.028	0.135	0.061
Calculus	0.039	0.014	0.174	0.072
Analytic geometry of space	0.024	0.034	0.119	0.038
Linear algebra	0.031	0.074	0.152	0.085
	Basic Discipline	**Engineering Major**	**Science Major**	**Basic Subject**
Advanced mathematics	0.051	0.063	0.047	0.027
Science and engineering	0.169	0.168	0.169	0.142
Required course	0.027	0.035	0.035	0.074
Math class	0.135	0.329	0.427	0.139
Series of numbers	0.164	0.062	0.174	0.027
Limits	1.241	0.048	0.538	0.157
Calculus	0.528	0.369	1.697	0.113
Analytic geometry of space	0.041	0.127	0.152	0.169
Linear algebra	0.274	0.035	0.196	0.044

Table 2. Calculation results of word similarity and weight

Keyword in Text *x*	**Advanced Mathematics**	**Calculus**	**Algebra**	**Geometry**
Maximum similarity	1.162	0.015	0.963	0.058
Weight	0.132	0.158	0.174	0.162
Corresponding keyword in *b* when the similarity in *a* is the largest *a*	Advanced mathematics	Calculus	Linear algebra	Analytic geometry of space
Weight of keywords in *b*	0.162	0.174	0.159	0.135
Keyword in Text *x*	**Basic Discipline**	**Engineering Major**	**Science Major**	**Basic Subject**
Maximum similarity	1.205	0.369	1.472	0.128
Weight	0.139	0.152	0.106	0.159
Corresponding keyword in *b* when the similarity in *a* is the largest *a*	Required course	Science and engineering	Science and engineering	Math class
Weight of keywords in *b*	0.162	0.185	0.169	0.126

Table 2 shows the calculation results of similarity and weights of words in text *a*. The similarity value of teaching content text and topic discussion text was 0.071445, which was taken as the reference value for rating the online learning sentiments of students.

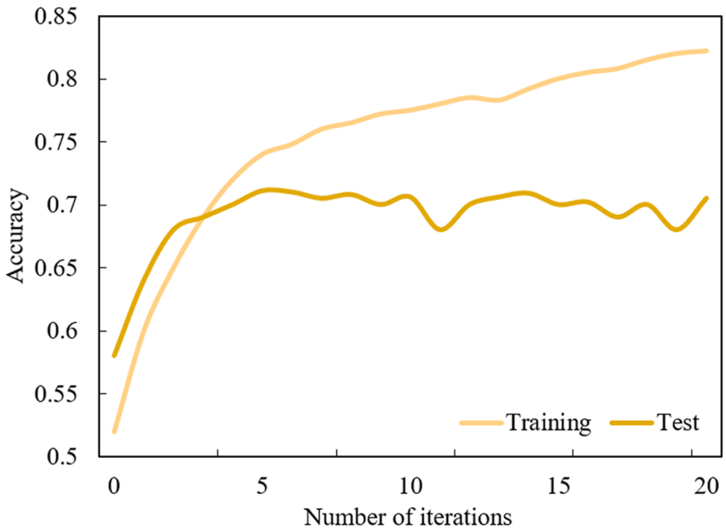

Fig. 6. Classification accuracy of the combined model

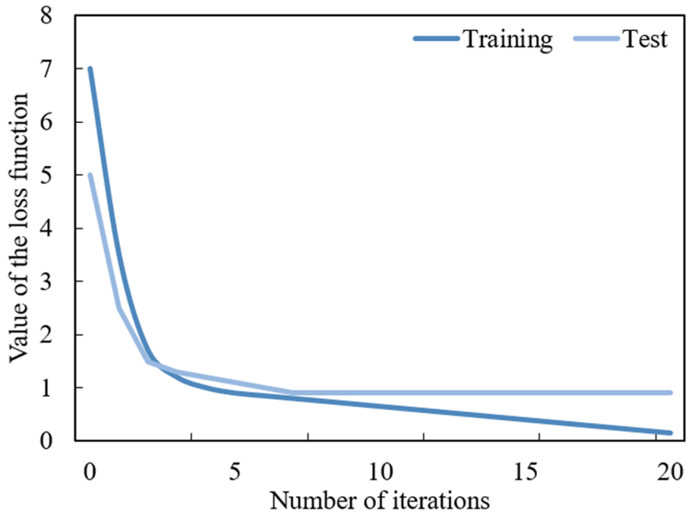

Fig. 7. Loss value of the combined model

To verify the effectiveness of the proposed model in classifying the learning sentiments of students, a comparative experiment was set to train and verify the proposed combined method and some reference models. The reference models include the conventional CNN, the conventional LSTM network, the proposed model before introducing the self-attention layer, and the *Word2vec-SVM* model. Figure 6 shows the classification accuracy of the combined model, and Figure 7 shows the loss of the combined model. As can be seen from the figures, the constructed combined model gradually converged after 20 iterations on the collected training sample set.

After training, the proposed combined model and other reference models were applied to the test sample set for testing, the same methods were adopted to train the other four comparative models on the dataset and make predictions, and the results of the comparative experiment are given in Table 3.

Table 3. Classification performance of each model on different sample sets

	Model	Accuracy	F_1 Value
Sample set 1	CNN	72.69	73.15
	LSTM	70.35	74.59
	Before introducing self-attention layer	76.29	73.62
	Word2vec-SVM	75.03	77.58
	The proposed model	81.82	89.51
Sample set 2	CNN	63.25	61.28
	LSTM	68.53	63.05
	Before introducing self-attention layer	72.36	73.69
	Word2vec-SVM	78.42	78.41
	The proposed model	80.29	86.14

According to the table, compared with other reference models, the proposed model outperformed other models in terms of the classification effect of students' learning sentiments on different sample sets. The main reason is that the combined model constructed in this paper had set two embedding layers: the static word embedding layer and the non-static word embedding layer, which had made full use of the high-quality word vector files, realized accurate capture of the information of students' learning sentiments, and attained more precise descriptions and representations of the sematic information of topic discussion.

5 Conclusion

This paper studied the classification of students' learning sentiments based on the mined data of the topic discussion texts of online learning platforms. At first, this paper gave the overall structure of the proposed LSC algorithm, and discussed the similarity between the topic discussion content and the teaching content. Then, it combined CNN with LST to build the said LSC model, so as to merge the advantages of the two and improve the accuracy of learning sentiment rating. After that, embedding layers of static words and non-static words were introduced into the proposed model for the purpose of realizing the mining of specific textual information while enhancing the semantic expression ability of the words. Combining with experiments, this paper gave the calculation results of sematic similarity and word similarity of the feature word sets and their weights, and took the similarity value of the teaching content and topic discussion content as the reference value for rating the online learning sentiments of students. Moreover, comparative experiment was set to verify the effectiveness of the proposed model in classifying the learning sentiments of students, and the classification accuracy and loss value of the combined model were given as well. At last, the classification performance of each model on different sample sets was attained, which had further verified the effectiveness of the proposed model.

6 References

[1] Yaseen, H., Alsoud, A.R., Nofal, M., Abdeljaber, O., Al-Adwan, A.S. (2021). The effects of online learning on students' performance: A comparison between UK and Jordanian universities. International Journal of Emerging Technologies in Learning, 16(20): 4–18. https://doi.org/10.3991/ijet.v16i20.24131

[2] Liu, Z., Kong, W., Liu, S., Kong, X. (2021). Using semi-supervised deep learning for identifying cognitive engagement in online learning discussion. In 2021 Tenth International Conference of Educational Innovation through Technology (EITT), Chongqing, China, pp. 233–238. https://doi.org/10.1109/EITT53287.2021.00053

[3] Kaoud, H., El-Shihy, D. (2021). Online learning in Egyptian universities post COVID-19 pandemic: A student's perspective. International Journal of Emerging Technologies in Learning, 16(18): 38–52. https://doi.org/10.3991/ijet.v16i18.25135

[4] Li Linbo, Wang Yanli, Wu Bing. Discussion on online classroom teaching mode under the concept of active learning[C]. Journal of Physics: Conference Series, July 28, 2021,v1976,n1,2021 International Conference on Computer, Communication, Control, Automation and Robotics, CCCAR 2021. https://doi.org/10.1088/1742-6596/1976/1/012070

[5] Aldossrey, S., Ali, A., Alalwani, R. (2020). The impact of group discussion in online teaching to enhance student learning in some Arabic and Islamic courses. In 2020 Sixth International Conference on e-Learning (ECONF), Sakheer, Bahrain, pp. 340–344. https://doi.org/10.1109/econf51404.2020.9385519
[6] Ali, M.F., Tahir, L.M., Said, M.N.H.M., Junaidi, J., Atan, N.A., Hahsan, A. (2017). Promoting staged self-directed learning (SSDL) among Malaysian tertiary learners through online discussion in completing group assignment. In 2017 IEEE 6th International Conference on Teaching, Assessment, and Learning for Engineering (TALE), pp. 265–270. https://doi.org/10.1109/TALE.2017.8252345
[7] Ouyang, F., Scharber, C. (2017). The influences of an experienced instructor's discussion design and facilitation on an online learning community development: A social network analysis study. The Internet and Higher Education, 35: 34–47. https://doi.org/10.1016/j.iheduc.2017.07.002
[8] Nguyen, N.L., Wang, M.H., Dow, C.R. (2021). Learning to recognize sockpuppets in online political discussions. IEEE Systems Journal, 16(2): 1873–1884. https://doi.org/10.1109/JSYST.2021.3117815
[9] Kim, D., Park, Y., Yoon, M., Jo, I.H. (2016). Toward evidence-based learning analytics: Using proxy variables to improve asynchronous online discussion environments. The Internet and Higher Education, 30: 30–43. https://doi.org/10.1016/j.iheduc.2016.03.002
[10] Maryuningsih, Y., Hidayat, T., Riandi, R., Rustaman, N.Y. (2019). Critical thinking skills of prospective biology teacher on the chromosomal basic of inheritance learning through online discussion forums. In Journal of Physics: Conference Series, 1157(2): 022090. https://doi.org/10.1088/1742-6596/1157/2/022090
[11] Xu, J., Sun, Y., Li, W. (2021). Organizational strategies of online discussion to promote college students' deep learning under the background of COVID-19. In 2021 4th International Conference on Education Technology Management, Tokyo, Japan, pp. 125–130. https://doi.org/10.1145/3510309.3510329
[12] Xu, B., Chen, N.S., Chen, G. (2020). Effects of teacher role on student engagement in WeChat-Based online discussion learning. Computers & Education, 157: 103956. https://doi.org/10.1016/j.compedu.2020.103956
[13] Pillutla, V.S., Tawfik, A.A., Giabbanelli, P.J. (2020). Detecting the depth and progression of learning in massive open online courses by mining discussion data. Technology, Knowledge and Learning, 25(4): 881–898. https://doi.org/10.1007/s10758-020-09434-w
[14] Hernández-Lara, A.B., Serradell-López, E. (2018). Student interactions in online discussion forums: their perception on learning with business simulation games. Behaviour & Information Technology, 37(4): 419–429. https://doi.org/10.1080/0144929X.2018.1441326
[15] Liu, Y., Jiang, C., Zhao, H. (2018). Using contextual features and multi-view ensemble learning in product defect identification from online discussion forums. Decision Support Systems, 105: 1–12. https://doi.org/10.1016/j.dss.2017.10.009
[16] Wang, H.C., Zhao, Y.L. (2020, November). Building an online learning question map through mining discussion content. In International Conference on Innovative Technologies and Learning, Porto, Portugal, pp. 367–372. https://doi.org/10.1007/978-3-030-63885-6_41
[17] Sheng, L. (2020). Discussion on the feasible application of "Online"+"Offline" mixed learning models of English in higher vocational colleges. In 2020 International Conference on Modern Education and Information Management (ICMEIM), Dalian, China, pp. 653–656. https://doi.org/10.1109/ICMEIM51375.2020.00147
[18] Nikolay, K., Alexander, D. (2019). Using deep learning to predict user behavior in the online discussion. In International Conference on Digital Transformation and Global Society, St. Petersburg, Russia, pp. 367–377. https://doi.org/10.1007/978-3-030-37858-5_30

[19] Chen, B., Chang, Y.H., Ouyang, F., Zhou, W. (2018). Fostering student engagement in online discussion through social learning analytics. The Internet and Higher Education, 37: 21–30. https://doi.org/10.1016/j.iheduc.2017.12.002
[20] Wu, S.Y., Hou, H.T., Shih, M.J.A. (2016). Exploring cognitive processing and behavioural patterns in a seamless online learning integrated discussion environment: a progressive sequential analysis. International Journal of Web Based Communities, 12(2): 195–211. https://doi.org/10.1504/IJWBC.2016.077255
[21] Jiao, X., Wang, R., Cai, S. (2022). Exploring the argumentation characteristics in SSI video discussion area of online video learning website. In 2022 10th International Conference on Information and Education Technology (ICIET), Matsue, Japan, pp. 166–170. https://doi.org/10.1109/ICIET55102.2022.9778977
[22] Patni, L.D.P., Sudiarta, I.G.P., Suparta, I.N. (2020). Development content and online discussion strategy in blended learning to improve student mathematical learning outcomes. In Journal of Physics: Conference Series, 1503(1): 012020. https://doi.org/10.1088/1742-6596/1503/1/012020
[23] Li, M., Chen, Y., Luo, H. (2020). Effects of grouping strategies on asynchronous online discussion: evidence from learning analytics and social network analysis. In 2020 International Symposium on Educational Technology (ISET), Bangkok, Thailand, pp. 273–276. https://doi.org/10.1109/ISET49818.2020.00066
[24] Barman, L., Engquist, M., Jansson, M., Enoksson, F. (2019). Interactions in online discussion forums-an underutilized resource for learning? In 2019 IEEE Frontiers in Education Conference (FIE), Covington, KY, USA, pp. 1–5. https://doi.org/10.1109/FIE43999.2019.9028512

7 Author

Mingzhe Niu, female, now a doctoral student and associated professor at the Innovation and Entrepreneurship school of the Zibo Vocational College of China, chief of the Innovation and Entrepreneurship Teaching Research Office, and director of the Entrepreneurship Research Institute; dedicated in the research of innovation and entrepreneurship education, education management, and international education of Chinese language for more than 20 years; writer of more than 10 published papers and 2 monographs, and host of 5 municipal-level and above scientific research programs.

Article submitted 2022-09-10. Resubmitted 2022-10-25. Final acceptance 2022-10-26. Final version published as submitted by the authors.