International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-16

https://doi.org/10.70917/ijcisim-2025-0231
Article

Analyzing Students' Emotional Tendencies in the Context of Civic Education by Combining the Parsimonious Bayesian Approach

Xiaoying Liu [1] and Xianyi Li [2,*]

[1] College of Fashion & Design, Zhejiang Fashion Institute of Technology, Ningbo, Zhejiang, 315000, China
[2] School of Mechatronics and Energy Engineering, NingboTech University, Ningbo, Zhejiang, 315000, China; lxy0102@163.com

Abstract: The study applies a plain Bayesian algorithm classification model with LDA topic model and feature-weighted fusion algorithm for the analysis of students' sentiment tendency in the context of different Civic and Political Education topics. Firstly, the text of Civic and political education is mined through word splitting, feature selection, weight replication, and combining the TF-IDF algorithm with the LDA model; secondly, a feature-weighted fusion of the plain Bayesian sentiment classification algorithm is proposed to realize the optimization of the sentiment classification, comparing the keyword TF and IDF values to refine the textual information, and to deepen the students' understanding of the educational content; the use of the LDA model is applied in the theme of the Civic and political education content mining, and then analyze the student comment data for sentiment tendency. The results show that the algorithm proposed in this paper is better than other sentiment classification algorithms, with an accuracy of 95.92% on two-category data and 82.76% on multi-category data. Eight topic types were extracted using LDA model. Finally, the students' affective tendency was calculated, and the affective quantity and affective value of the affective features and themes in each stage of different contexts were output, which verified the accuracy and validity of the model's classification.

Keywords: TF-IDF; LDA theme model; civics course; theme extraction

1. Introduction

“Intelligent era, what is education” is a major proposition of the current era, the new era, colleges and universities should closely focus on strengthening the core of ideological and political education, take the initiative to embrace artificial intelligence (AI) and other technologies, to explore the depth of its integration with the ideological and political education of the new model, which is not only the new era of colleges and universities to practice This is not only the inevitable requirement of the mission of “educating people for the Party and educating talents for the country”, but also a key initiative to improve the effectiveness of ideological and political work and enhance the sense of the times and attractiveness [1-4]. The traditional model of Civic and political education can no longer meet the needs of students' precision and personalization, and the educational integration model of AI+Civic and political education makes it possible to cultivate precise and personalized talents, in which the analysis of students' emotional tendency plays an important role [5-7].

For the analysis of students' affective tendencies, researchers have made use of online education data, online social data, online learning interaction data, etc. Wang et al. [8] constructed a semantic analysis model for analyzing the affective tendencies of students in catechism classes, predicting the probability of graduation at a certain stage of the course with the support of affective quantification and machine learning, providing personalized guidance for students with specific affective tendencies, and promoting graduation. Deng [9] used the implicit Dirichlet distribution model to build a joint model of thematic emotions, and introduced a long and short-term memory network algorithm to portray students' network portraits, which was used to analyze the emotional tendencies of college students' networks. Moreno-Ger and Verdú [10] developed a ubiquitous emotion model for the virtual classroom, which can record students' facial expressions, and form the emotions of the students in the classroom into an emotional cluster, based on which students' emotional tendencies are inferred. Zhang et al [11] analyzed students' emotional tendencies in the classroom using a random forest model, and found that students' emotional tendencies were affected by various aspects of the teacher and the students, such as the time of the students' speeches and the teacher's motivation.

Currently, fewer studies analyze the emotional tendency of students in Civics education, and by analyzing and tracking the evolution of students' emotional tendency, it has an important practical value for improving the effect of Civics education. Yuzhong [12] combined AI and data mining to analyze the emotion of students in Civics classroom, which included students' own emotion and emotional response under external stimuli, and explored the correlation between emotion analysis the correlation between emotion analysis and the effectiveness of Civics teaching. Qiujiang [13] used deep learning models to analyze the emotional tendencies under students' comments in the feedback system of Civics education resources, and performed quantitative analysis of emotions, so that the system could improve the system resources with more students' emotions and assist teachers in making personalized teaching plans.

The Plain Bayes (NB) method is a classification method based on Bayes' theorem and the assumption of conditional independence of features. Bayes theorem it can effectively deal with the category judgment problem, after considering the conditional independence assumption, the plain Bayesian method greatly simplifies the computation and can solve complex practical problems, it also has a variety of advantages such as high efficiency, noise resistance, etc., and exhibits a better categorization effect, which shows a strong potential in the analysis of sentiment tendency [14-16]. And Abbas et al [17] used polynomial NB to classify the comments of a website, so as to obtain the sentiment vocabulary therein for classification and realize sentiment analysis, and its accuracy is better than that of the ordinary NB method. Li et al [18] combined the NB model and the sentiment lexicon to do the sentiment analysis of pop-ups in the video, which can be classified into 7 dimensions, and the introduction of weighting calculation can also be used to polarize the sentiment for Classification.

In order to explore the emotional state of students in the context of Civic Education, this paper utilizes the Plain Bayesian method. The TF-IDF algorithm is used to process word frequency information effectively, and then the feature-weighted fusion of plain Bayesian sentiment classification algorithm is proposed, and experiments are designed to verify the effectiveness of the algorithm. Then through word frequency statistics, TF-IDF value calculation, and combined with the LDA model to accurately determine the topic text of Civic Education, laying the foundation for the subsequent sentiment analysis. Then, from the perspective of sentiment analysis, the degree of emotional tendency was calculated, and the feature-weighted fusion of plain Bayes was used to characterize the sentiment analysis of the preprocessed text and LDA-extracted themes, and the number of sentiment features and the sentiment value of the themes in each stage of different contexts were outputted from the three stages of latency, diffusion, and recession.

2. Theme mining of Civic Education Texts based on LDA Modeling

2.1. Civic Education Text Mining Process

2.1.1. Segmentation

Based on the understanding of mainstream text mining techniques, the text mining steps taken in this paper are shown in Figure 1.

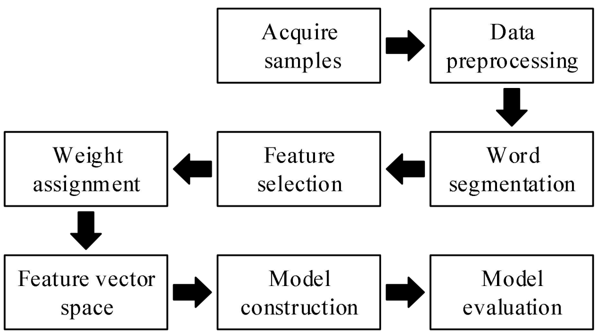


Figure 1. Text mining steps.

Segmentation, i.e., splitting the words in a text, transforms the text into a sequential arrangement of multiple words. Segmentation is equivalent to feature extraction, and the extracted words are called feature words. The general methods of word separation are: dictionary-based methods; annotation-based methods; rule-based methods; statistical-based methods.

2.1.2. Feature selection

Feature selection, i.e., after obtaining the feature words, the vector space dimension is too large due to the more complex text and larger word content. For some prepositions, commonly used words are not considered, these screening deleted. Generally deleted through the deactivation thesaurus, commonly used Chinese thesaurus: HIT deactivation thesaurus, Sichuan University Machine Learning Intelligence Laboratory deactivation thesaurus, Baidu deactivation list, etc., this paper adopts the deactivation thesaurus of Sogou.

After removing the deactivated words, the feature words that play a larger role in text classification and topic analysis are selected and arranged in order of importance. The feature selection methods include: document frequency (DF), mutual information (MI), information gain (IG), χ^2 -statistic (CHI), textual evidence weight (WET), expected cross entropy (ECE), odds ratio (OR), and word weight (TS). Typically, words that occur in all texts are less important and are screened out in the feature selection session, leaving behind words with more distinctive features.

2.1.3. Assignment of weights

Weight assignment, which is done to optimize the text space, is used to normalize the text vectors by using the most appropriate computational methods so that they can be better used for text classification and text clustering, etc. The weight value can be added to each feature word in each text, and the weight assignment can be performed on each feature word on each text. The weight assignment method used in this paper is TF-IDF, which will be introduced in detail later.

After completing the feature vector space transformation of text, the purpose of text classification or text clustering is achieved. In statistics and machine learning, many algorithms on classification and clustering have been developed, and most of them have been gradually matured and put into a large number of practical application scenarios with good results.

In the application scenario of text analysis, the common algorithms include LDA topic model for topic analysis; common algorithms include Simple Bayesian Classification, Support Vector Machine (SVM), Convolutional Neural Networks for text classification; and common algorithms include K-means algorithm, Knn algorithm for text clustering, etc. In this paper, we will use LDA topic model to do the classification and clustering. In this paper, LDA topic model will be used for further analysis.

By constructing the model and inputting the tested text into the model, the evaluation results of the model will be finally obtained according to the model output results compared with the actual results.

2.2. TF-IDF algorithm

The purpose of the TF-IDF algorithm [19] is to assess the importance of a word to a text, and a word or phrase is considered to have good category differentiation if it occurs more frequently in a text and less frequently in a document set. The TF-IDF consists of two components: word frequency and inverse document frequency.

Word frequency is the frequency of occurrence of a specific word in a document:

$$\text{TF} = \frac{\text{The number of times a specific word appears in the article}}{\text{The total number of words in the article}} \tag{1}$$

The inverse document frequency is the logarithm of the ratio of the total number of documents to the number of documents containing a specific word:

$$\text{IDF} = \log\left(\frac{\text{Total number of documents in the corpus}}{\text{Number of documents containing the word + 1}}\right) \tag{2}$$

If the word frequency of occurrence of a keyword i in a text d is $tf_i(d)$ and n_i is the number of texts containing the keyword i, the TF-IDF function is:

$$\text{TF-IDF}_i(d) = tf_i(d) \cdot \log\left(\frac{N}{N_i + 1}\right) \tag{3}$$

Therefore, in this paper, the TF-IDF algorithm can extract and calculate the keywords and weights in a certain text.

2.3. LDA Subject Modeling

LDA Latent Dirichlet Distribution Topic Modeling is an unsupervised machine learning technique that can be used to identify and generate classifications of latent topics in large-scale document collections or corpora [20]. It adopts the bag-of-words approach, by constructing a three-layer Bayesian probabilistic model of "document-topic-word", it can obtain the distribution probability of words under each topic and the corresponding topic probability of the document, so as to categorize the topics according to the obtained probabilities.

In the LDA model, the generation process of a document is as follows:

1) The topic distribution θ_i corresponding to document d_i is generated by sampling from the Dirichlet distribution α .

2) The topic $Z_{i,j}$ of the j th word of document i is generated by sampling from the polynomial distribution of topics θ_i .

3) The corresponding word distribution $\phi_{Z_{i,j}}$ for topic $Z_{i,j}$ is generated by sampling from the Dirichlet distribution β .

4) Synthesize the polynomial distribution of words $\phi_{Z_{i,j}}$ corresponding to the topic $Z_{i,j}$ to generate the words $W_{i,j}$.

Where the parameters α , β and the number of topics corresponding to the theme K are generally given in advance, the graph indicates the dependencies as vector edges, the repetitions as rectangles, and the repetitions as letters M , N , K , and the number of times the repetitions are indicated in the rectangles. Therefore, to generate a document, the conditional probability formula for the occurrence of each word in the document is as follows:

$$P(w \mid d) = P(w \mid z) \cdot P(z \mid d) \tag{4}$$

The probability of occurrence of each word in each document is $P(w \mid d)$; the frequency of occurrence of each word in each topic is $P(w \mid z)$; the probability of occurrence of each topic in each document is $P(z \mid d)$. When given a text set, each text in the text set can be divided into words, and the word frequency of each word in each text can be calculated to obtain the "document-word" matrix. The LDA topic model is to infer the topic of a document with a certain probability through the three layers of "document-topic-word".

3. Algorithm for classifying students' emotions in the contextual themes of Civic Education texts

3.1. Quantitative Representation of Civic Education Texts

The text words are converted into word vectors using Word2vec technique, and the CBOW model is used to predict using contextual information w_{t-2} , w_{t-1} , w_{t+1} , w_{t+2} . The input layer inputs the word vectors of the context and the output layer calculates the word vectors with the highest probability, the model is shown in Fig. 2. The Gensim package for python is utilized for word vector extraction.

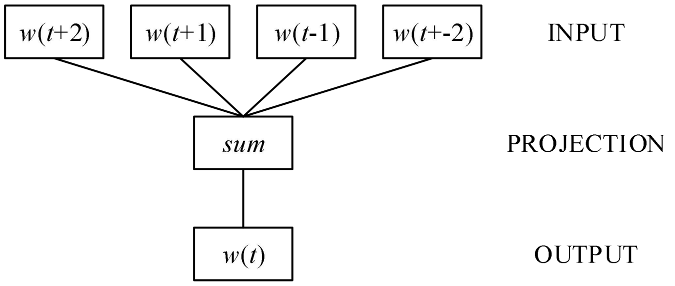


Figure 2. CBOW illustraton of model.

3.2. Feature-weighted fusion of plain Bayesian sentiment classification algorithms

3.2.1. Feature extraction

Feature extraction is the basis for realizing the judgment of emotional tendency of text. The effect of the classification method using emotion words as feature words is better than the classification effect using CHI feature selection method. Therefore, in this paper, an extended emotion dictionary is obtained by integrating the emotion dictionaries of texts under different topics of Civic and Political Education as well as the collected emotion words, and filtering out the effective emotion words to be added to the emotion word collection. Finally, feature extraction is performed by combining the constructed extended emotion lexicon.

3.2.2. Bayesian classification models

For text $d = \{w_1, w_2, \ldots, w_n\}$ the sentiment tendency belongs to $j = \{\text{pos}, \text{neg}\}$, assuming that in the case of the individual sentiment features are independent of each other, the Plain Bayesian Sentiment Classification formula is obtained:

$$P(C_j \mid d) = \arg\max p(c_j)\prod_{i=1}^{n} p(w_i \mid c_i) \tag{5}$$

where $p(c_j)$ is the prior probability that the text category is j; $p(w_i \mid c_i)$ is the posterior probability of the Bayesian model.

$$p(c_j) = \frac{\sum_{i=1}^{n} s(w_i, c_j)}{|N|} \tag{6}$$

$$p(w_i \mid c_j) = \frac{s(w_i, c_j)}{\sum_{i=1}^{n} s(w_i, c_j)} \tag{7}$$

where $\sum_{i=1}^{n} s(w_i, c_j)$ is the total number of sentiment words contained in documents with category j; $|N|$ is the total number of sentiment words contained in the entire training corpus of documents; and $s(w_i, c_j)$ is the sum of the number of times the sentiment word w_i has appeared in each document in the context of a sentiment category of j. .

In addition, in order to avoid the case that the posterior probability $p(w_i \mid c_j)$ is equal to 0 during the computation process, the Laplace transformation is used to finally obtain the formula for calculating the posterior probability:

$$p(w_i \mid c_j) = \frac{s(w_i, c_j) + 1}{\sum_{i=1}^{n} s(w_i, c_j) + |V|} \tag{8}$$

where $|V|$ is how many sentiment words the training sample contains that are not repeated.

3.2.3. Emotion word feature weighting

Therefore, in this paper, we determine the degree of contribution of sentiment words of different polarity to the analysis of text sentiment polarity by introducing a weighting factor $T(w_i \mid c_j)$. $T(w_i \mid c_j)$ denotes the weight of the sentiment word w_i in the comment text judged to be of category j. The larger the value of the weighting factor, the greater the degree of contribution of the emotion word to the overall emotional polarity analysis of the text, i.e., the greater the weight of the emotion word's contribution to the overall emotional polarity of the text. After analysis, it is found that the weight value of the positive emotion word's contribution to the positive comment text should be greater than the weight value of the positive emotion word's contribution to the negative comment text. Similarly, the weight value contributed by negative sentiment words to negative comment text should be greater than the weight value contributed by negative sentiment words to positive comment text. Therefore, in order to more effectively assign the corresponding weights to the emotion words, this paper determines the formula of the weighting factor by combining the number of emotion words in the test text. Let p denote the number of positive sentiment words in the test text, and q denote the number of negative sentiment words in the first test text. In this paper, the formula for calculating the weighting factor is defined as conforming to the following equation:

$$T(w_i \mid c_j) = \begin{cases} 2\times(p+q), & \text{If } w_i \in L_{\text{pos}} \text{ And } j = \text{pos} \\ p+1, & \text{If } w_i \in L_{\text{pos}} \text{ And } j = \text{neg} \\ q, & \text{If } w_i \in L_{\text{neg}} \text{ And } j = \text{pos} \\ q+p, & \text{If } w_i \in L_{\text{neg}} \text{ And } j = \text{neg} \end{cases} \quad p > q \tag{9}$$

$$T(w_i \mid c_j) = \begin{cases} p+q, & \text{If } w_i \in L_{\text{pos}} \text{ And } j = \text{pos} \\ p, & \text{If } w_i \in L_{\text{pos}} \text{ And } j = \text{neg} \\ q+1, & \text{If } w_i \in L_{\text{neg}} \text{ And } j = \text{pos} \\ 2\times(q+p), & \text{If } w_i \in L_{\text{neg}} \text{ And } j = \text{neg} \end{cases} \quad p < q \tag{10}$$

$$T(w_i \mid c_j) = \begin{cases} p+q+1, & \text{If } w_i \in L_{\text{pos}} \text{ And } j = \text{pos} \\ p+1, & \text{If } w_i \in L_{\text{pos}} \text{ And } j = \text{neg} \\ q, & \text{If } w_i \in L_{\text{neg}} \text{ And } j = \text{pos} \\ 2\times(q+p+1), & \text{If } w_i \in L_{\text{neg}} \text{ And } j = \text{neg} \end{cases} \quad p = q \tag{11}$$

where L_{pos} is the set of sentiment words belonging to the positive category in the constructed sentiment dictionary; L_{neg} is the set of sentiment words belonging to the negative category in the constructed sentiment dictionary.

3.2.4. Overall feature weighting

In order to more effectively assign corresponding weights to the features as a whole, this paper combines the following 2 conditions:

1) When the number of positive sentiment words appearing in each test text is greater than the number of negative sentiment words, the weights contributed by the positive sentiment words to the positive comment text are greater than the weights contributed by the negative sentiment words to the negative comment text;

2) When the number of negative sentiment words appearing in each test text is greater than the number of positive sentiment words, the weight of the negative sentiment words contributing to the negative comment text is greater than the weight of the positive sentiment words contributing to the positive comment text.

Combining the above 2 conditions, this paper determines the overall weight of the impact of the number of sentiment words on the text sentiment analysis by introducing a weighting factor. $w(w_+ \mid c_j)$ denotes the weight of the overall contribution of positive sentiment words to the comment text, and $n(w_- \mid c_j)$ denotes the weight of the overall contribution of negative sentiment words to the comment text. The formula for calculating the overall weighting factor is defined as:

$$w(w_+ \mid c_j) = \begin{cases} (p+1)\times(q+1), & \text{If } q < p \text{ And } j = \text{pos} \\ q+1, & \text{If } q < p \text{ And } j = \text{neg} \\ p+1, & \text{If } p < q \text{ And } j = \text{pos} \\ q, & \text{If } p < q \text{ And } j = \text{neg} \\ (q+1)\times(p+1), & \text{If } p = q \text{ And } j = \text{pos} \\ p+1, & \text{If } p = q \text{ And } j = \text{neg} \end{cases} \tag{12}$$

$$n(w_- \mid c_j) = \begin{cases} p, & \text{If } q < p \text{ And } j = \text{pos} \\ p+1, & \text{If } q < p \text{ And } j = \text{neg} \\ q+1, & \text{If } p < q \text{ And } j = \text{pos} \\ (q+1)\times(p+1), & \text{If } p < q \text{ And } j = \text{neg} \\ q, & \text{If } p = q \text{ And } j = \text{pos} \\ q+1, & \text{If } p = q \text{ And } j = \text{neg} \end{cases} \tag{13}$$

3.2.5. Feature Weighted Fusion Algorithm

In order to further improve the accuracy of text sentiment classification as well as to enhance the comprehensive performance of the classification model, this paper fuses the two weighting factors introduced above to realize a feature-weighted fusion of plain Bayesian sentiment classification algorithm. The fused formula is:

$$P(C_j \mid d) = p(c_j)w(w_+ \mid c_j)\cdot\prod_{i=1}^{n} p(w_i \mid c_j)\times T(w_i \mid c_j) \tag{14}$$

The feature-weighted fusion plain Bayesian sentiment classification algorithm is described as follows.

Input: training text, test text, ESL sentiment dictionary.

Output: sentiment tendency of the text.

Algorithm process:

1) Text preprocessing. Segment the text by the segmentation tool and combine with the deactivation word list to remove irrelevant words.

2) Feature extraction. Match the text with the constructed emotion dictionary to extract the corresponding emotion feature words.

3) Construction of classification model. Combine the feature words extracted in step 2 to train the training text with emotion labels.

4) Emotion word feature weighting. According to equations (9) to (11), calculate the value of the weighting factor $T(w_i \mid c_j)$ in different cases and combine it with $p(w_i \mid c_j)$ to adjust the posterior probability of the sentiment words in the Bayesian model.

5) Overall feature weighting. Calculate the weights of the overall weighting factors $w(w_+ \mid c_j)$ and $n(w_- \mid c_j)$ in different cases according to Eqs. (12)~(13), respectively, and incorporate them into the Bayesian classification model to adjust the overall probability of the model according to the number of sentiment words in the text.

6) Sentiment classification. Combining step 4) and step 5), the fused model is obtained, and the probability of the emotional category to which the text belongs is tested by calculating the probability of the emotional category to which the text belongs, and the one with a large probability is judged to be the emotional category to which the text belongs.

3.3. Experiments and Analysis of Sentiment Categorization

Comparing the accuracy, precision, recall, and $F1$ scores of sentiment dictionary-based methods, traditional machine learning-based methods (plain Bayes, random forest [21], logistic regression, and support vector machine), and feature-weighted fusion of plain Bayes sentiment classification algorithms, the dichotomous classification results are summarized as shown in Table 1 and Figure 3.

As a whole, it can be seen that the performance of the four classification models in traditional machine learning methods in terms of accuracy, for example, in this experimental dataset in order from largest to smallest, then in order of logistic regression model > support vector machine model > plain Bayesian model > random forest model. The accuracy of the plain Bayesian model is 3.62% higher than that of the random forest model, the accuracy of the support vector machine model is 0.75% higher than that of the plain Bayesian model, and the mean values of the evaluation metrics of accuracy, precision, recall, and $F1$ score for the four traditional machine learning classification methods are 0.8699, 0.8797, 0.8570 and 0.8669. While the classification method based on sentiment lexicon is less effective in it, the feature-weighted fusion of plain Bayesian sentiment classification method shows a good classification effect, with an accuracy rate of 95.92%, has a deeper feature learning ability, and is 8.93% more accurate than the average value of traditional machine learning classification methods.

Table 1. Summary of experimental results of various dichotomization methods.

Classification method	Accuracy	Precision	Recall	F_1 Score
ED	0.7748	0.7256	0.8824	0.7978
NB	0.8776	0.8933	0.8303	0.862
RF	0.8414	0.8907	0.7728	0.8289
LR	0.8811	0.9079	0.8443	0.8762
SVM	0.8851	0.8975	0.8546	0.8769
This method	0.9592	0.9632	0.9573	0.9598

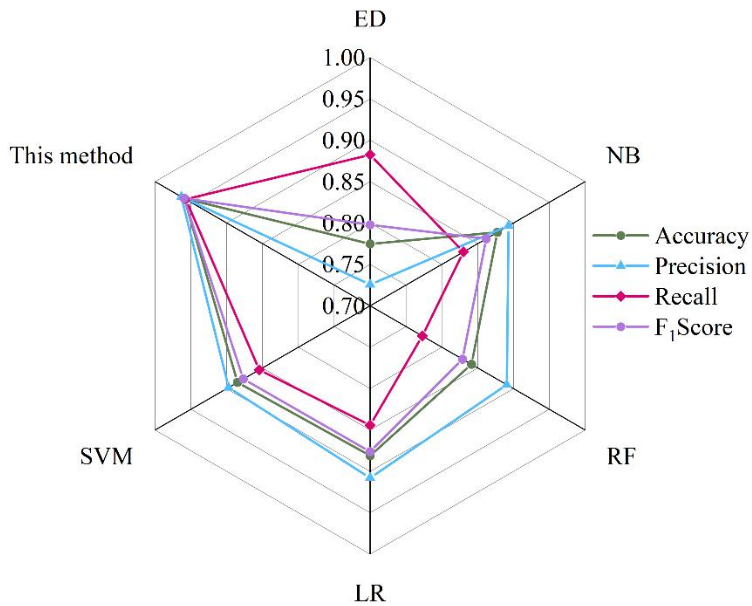


Figure 3. Binary classification method experimental radar chart.

Overall, the word vector in the machine learning model can capture the semantic information of the vocabulary, and the analysis process combined with the word vector can incorporate the semantic information into the comprehensive consideration, not only utilizing the sentiment vocabulary but also utilizing the information of other vocabulary, which can make the information of the textual data fully explored, while the sentiment dictionary is difficult to comprehensively cover all the vocabulary, and it can't be flexible and changeable for some vocabulary with different contexts and different meanings. Flexible changes, so according to the performance in this experimental dataset in descending order, then in order of this paper's method > traditional machine learning model > sentiment dictionary classification method. The three classification results are summarized as shown in Table 2 and Figure 4.

It can be seen that the performance of the four classification models in this experimental dataset in terms of accuracy in traditional machine learning methods is ranked from largest to smallest, then in order of logistic regression model > support vector machine model > plain Bayesian model > random forest model. The mean values of the evaluation metrics accuracy, precision, recall, and $F1$ score of the four traditional machine learning classifiers are 0.7640, 0.6880, 0.6922, and 0.6883, respectively.Overall, the feature-weighted fusion of the plain Bayesian sentiment classification method, on the other hand, exhibits a good classification effect, with a deeper feature learning capability.

Table 2. Summary of experimental results of three classification methods.

Classification method	Accuracy	Precision	Recall	F_1 Score
ED	0.5651	0.6027	0.6388	0.6148
NB	0.7983	0.6938	0.6931	0.6932
RF	0.7756	0.6632	0.6590	0.6587
LR	0.8130	0.7163	0.7152	0.7158
SVM	0.8101	0.7125	0.7108	0.7116
This method	0.8276	0.7392	0.7362	0.7354

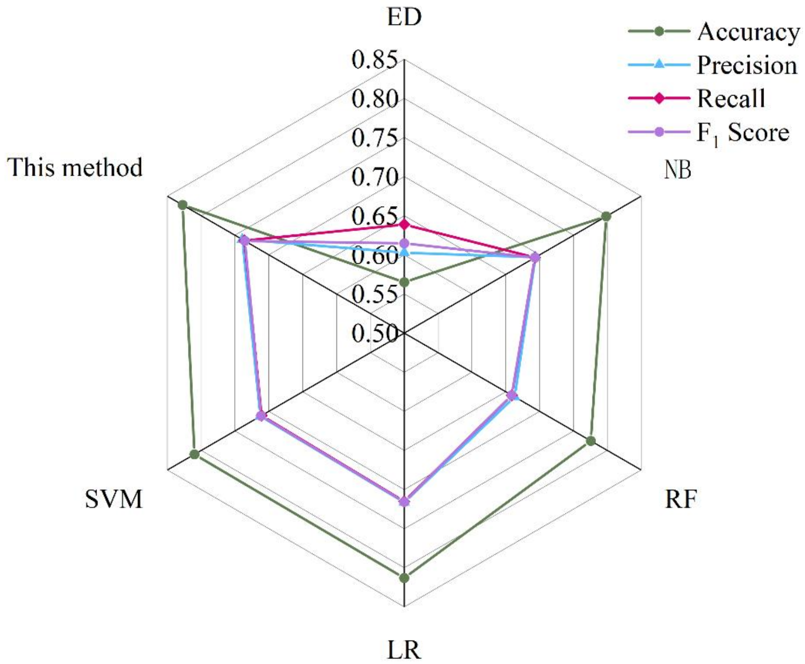


Figure 4. Three-category and multi-category method experimental radar chart.

4. Case Study of Students' Emotional Tendencies in the Context of Civic Education

Taking the text of ideological and political education documents from a teaching lecture on ideological and political education in a certain university as an example, the lecture mainly discussed the themes of "the spirit of joining the Party", "War and Revolution", "War History", "Self-awareness", "personal development", "National strength", "Comparison of National development", and "international situation". Explore the emotional tendencies of students under this text topic.

4.1. Word Frequency Analysis of Civic Education Texts

4.1.1. Word frequency statistics

Word frequency is the number of times a given word occurs in a document, which is used to assess the importance of a word in a document or corpus. Word frequency statistics, on the other hand, is the process or result of counting the frequency of occurrence of a word or phrase in a document or corpus, which provides a quantitative basis for text analysis.

The word frequency statistics of the Civic Education text when the lecturer was giving a lecture are shown in Figure 5, which shows that the lecturer put the country and the people in the first and second place in his speech with high frequency. In the first 11 high-frequency words of the pooled speech, the main emphasis is on the CPC's adherence to socialism, which highlights the central position of the CPC's leadership as well as its adherence to socialism, while the ultimate goal is to realize the great rejuvenation of the Chinese nation.

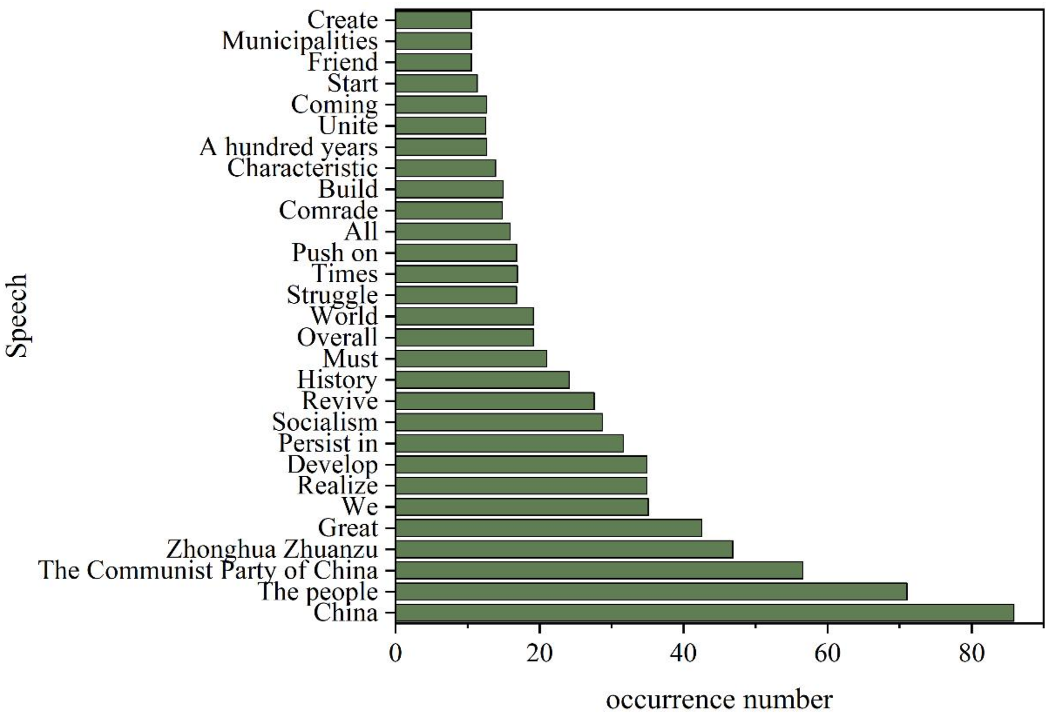


Figure 5. Speech frequency statistics.

4.1.2. Keywords TF-IDF value calculation

In theory or policy learning, uncommon and new words often have more information content, which requires the introduction of inverse document frequency IDF to accurately measure the weight of keywords. The IDF value of a word is generally calculated based on the corpus, and the more common it is in the corpus, the smaller its IDF value is.

The top 20 keywords in the TF-IDF value of the Civic Education text are shown in Table 3. The table shows that compared with the TF value of the keywords, the IDF value of the keywords reveals the change of the CPC's attention to governance, and the keywords in the top 10 of the IDF highlight the importance of "realizing the great rejuvenation of the Chinese nation", whereas the keywords in the top 10 of the TF basically only reflect the natural relationship between the state, the people, the political party and the nation, and can reflect the speech's "great revival of the Chinese nation". The first 10 keywords in IDF highlight the importance of "realizing the great rejuvenation of the Chinese nation", while the first 10 keywords in TF can only reflect the natural relationship between the state, the people, the political party, and the nation.

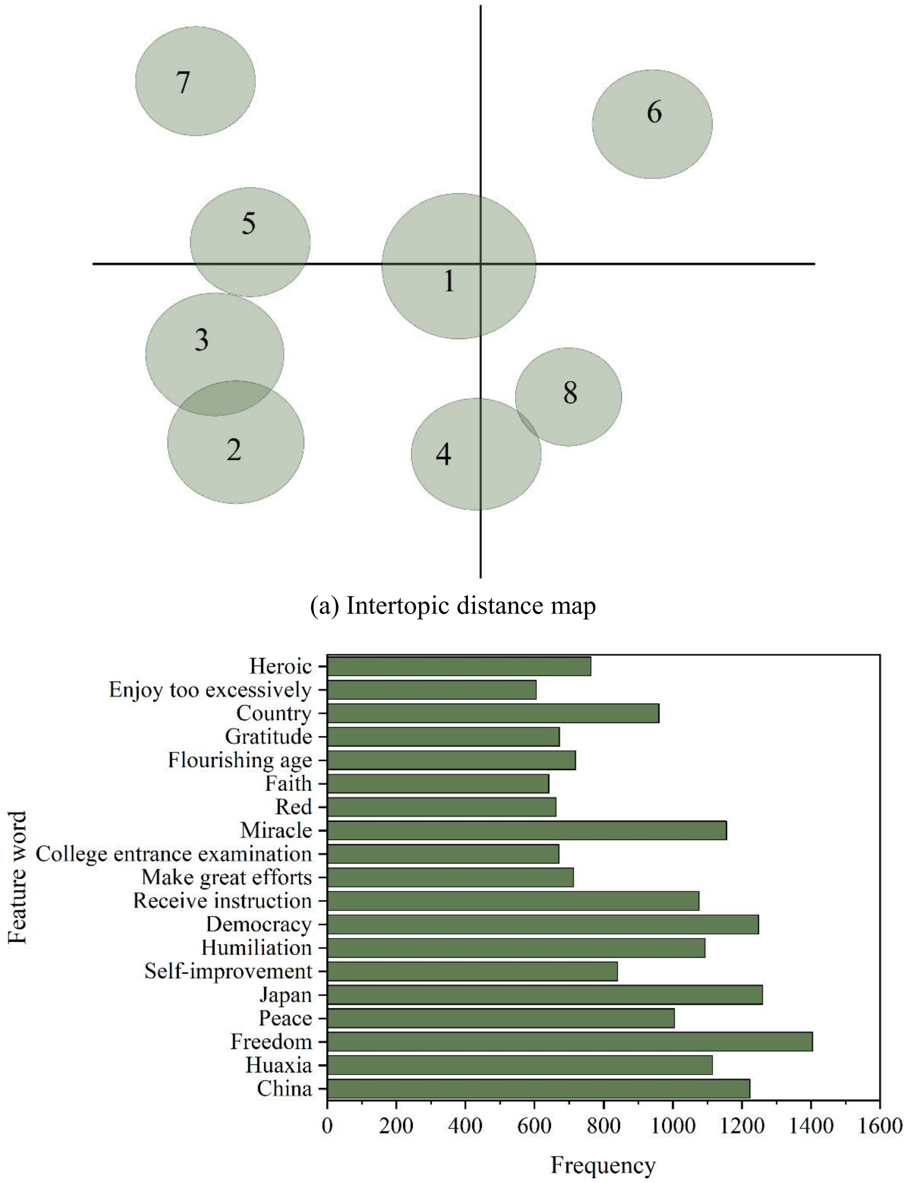
Table 3. Top 20 keywords for tf-df values.

Order number	High TF value keywords	High IDF keywords	Order number	High TF value keywords	High IDF keywords
1	China 84	Chinese nation 0.1574	11	Revival 28	Development 0.0589
2	People 69	Chinese Communist Party 0.1575	12	History 25	We 0.0571
3	The Communist Party of China 55	People 0.1558	13	Must 22	Learn from history 0.0568
4	The Chinese nation 46	Great 0.1208	14	Overall 20	Centenary 0.0541
5	Great 42	China 0.1121	15	World 20	History 0.0538
6	We 35	Revive 0.0976	16	Struggle 18	Overall 0.0529
7	Implement 35	Persist in 0.0843	17	Times 18	Journey 0.0526
8	Development 35	Socialism 0.0833	18	Push on 18	A hundred years 0.0509

9	Persist in 32	Achieve 0.0752	19	All 17	Must 0.0501
10	Socialism 29	Struggle 0.0649	20	Comrade 16	Push on 0.0489

4.2. Thematic Analysis of Civic Context Based on LDA Modeling

In the theme analysis stage, the data were de-weighted and cleaned to filter out the content discussed by students in the Civics lectures, which amounted to 121,024 entries. Using the dataset that has been processed by data cleaning, text segmentation, removal of deactivated words, and vectorization, the pre-trained completed model (min_sample = 6; $min\_\ cores$ = 4; ε = 0.3; the number of integrated models is 16) for thematic analysis to obtain the distribution of themes for each piece of data. Figure 6 shows the visualization results of theme clustering generated using pyLDAvis tool, after selecting a class of themes the distribution of related keywords for that theme can be seen, the association between themes and words is done by adjusting the size of the parameter λ, the distribution of words under the selected λ = 1 class of themes is the result of theme clustering.



(a) Intertopic distance map

(b) Top-19 most salient terms

Figure 6. LDA model topic clustering visualization results.

The above figure demonstrates the frequency of occurrence of feature words under each topic in all the data with the frequency of occurrence under the selected topics, and the frequency of occurrence of feature words within a topic is referable to the determination of the content of that topic. By comparing the distribution probabilities of the topic words in each case, the following eight topics that can describe the content of students' discussions in all the Civics contexts in the test set are formed, and the topic keywords are shown in Table 4.Topic6 and Topic7 demonstrate the development of changes in the national strength in the early days of the founding of New China and the 70th anniversary celebrations to arouse strong emotional resonance, and enhance the sense of national pride and Topic1 and Topic5 tell their own stories from the perspective of ordinary people, which convey positive values and outlook on life although they do not deliberately talk about theoretical knowledge. Through the innovative teaching form of Civics, we can enhance classroom participation, increase students' enthusiasm in learning, establish moral character, strengthen college students' belief in Marxism, cultivate their correct values of life, and build up their lofty ideals.

Table 4. LDA model topic clustering and topic word distribution.

Ttheme	Text keywords and content
Topic1	China, Long live China, Huaxia, Without regret
Topic2	America, Freedom, Peace, Angina
Topic3	Japan, Humiliation, Accept, Self-improvement
Topic4	Democracy, The west, Receive instruction, Mother country
Topic5	Make great efforts, Young people, College entrance examination
Topic6	Miracle, Red, Come on, Faith
Topic7	Flourishing age, Wish, Heroic, Gratitude
Topic8	Country, True, Esteem, Enjoy too excessively

4.3. Analysis of Students' Emotional Tendencies in the Context of Civic Education

4.3.1. Calculation of affective tendencies per unit of time

In this chapter, the sample data between November 1, 2025 - November 18, 2025 were computed on a day-by-day basis, and the average sentiment for each statistical time period was computed using a feature-weighted fusion of the plain Bayesian sentiment classification algorithm. When a lecture on receiving Civics education occurred, students influenced by different Civics topics started to post their opinions and attitudes about the lecture on the school forum one after another. The overall average emotion value, positive average emotion value, negative average emotion value and emotion heat over time within the unit are shown in Figures 7 to 9. Since only Topics 2 and 3 of the eight themes have heavier content, involving war scenes and overall emotions that cause a large number of emotional resonance, the overall emotion value is dominated by negativity.

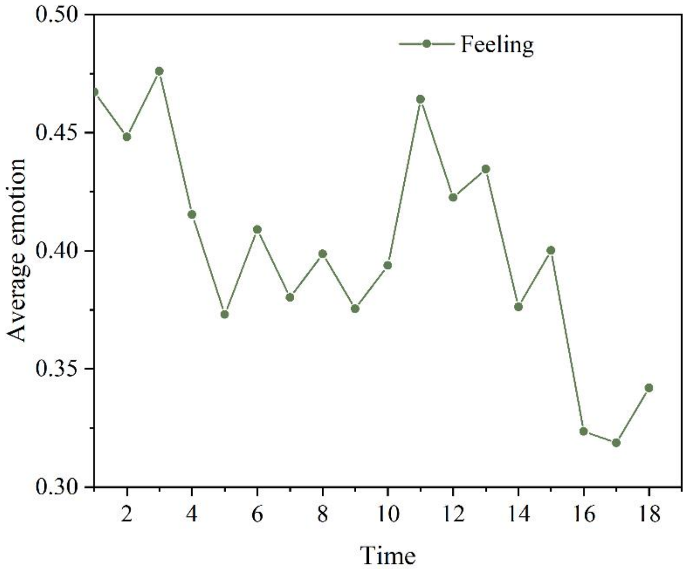
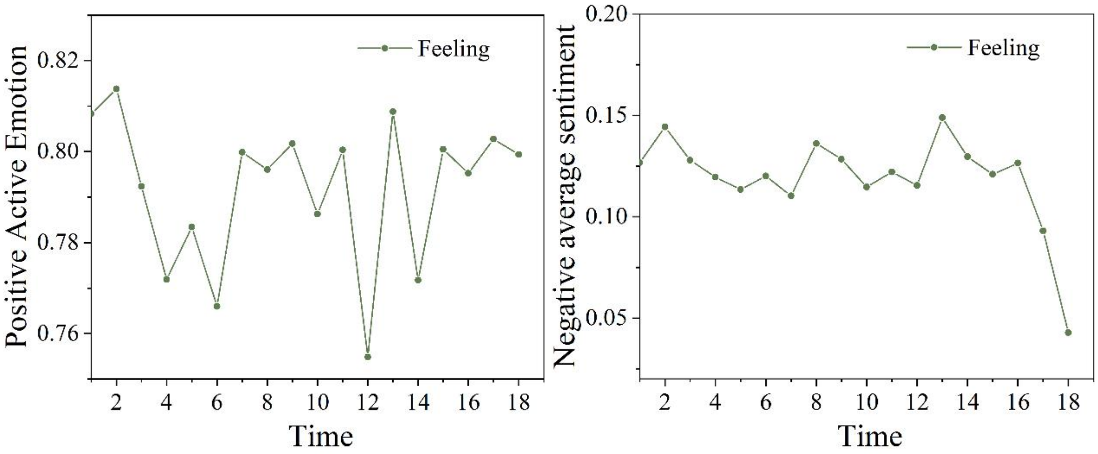


Figure 7. The overall average emotional value per unit of time.



(a) Positive Active Emotion (b) Negative average sentiment

Figure 8. The average positive and negative emotions per unit of time.

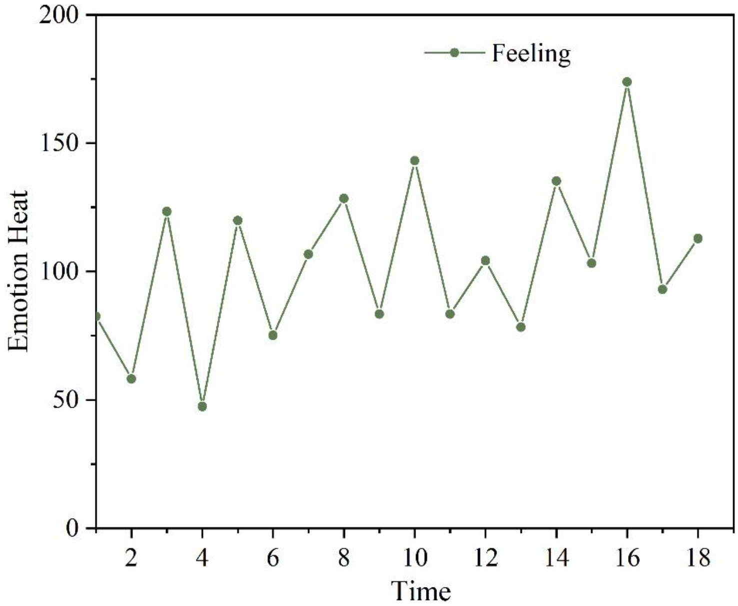


Figure 9. Emotion heat map over time.

4.3.2. Characterization of Emotional Characteristics by Stage Based on Improved NB

In this chapter, the trained improved plain Bayesian model is tested using performance measures, and the results show that the trained model has a good classification effect, so based on the three-stage theory, the number of positive, neutral and negative emotions and the value of emotions in different stages are counted, as shown in Table 5.

As can be seen from the above table, the distribution of the three types of emotions of students in each stage is relatively average, and the gap is not very large. In the latent period of public opinion, students' positive emotions accounted for 28.90%, and the emotional value was 0.7827, mainly because at the early stage of Topic1, due to the asymmetry of information, students' perception of the event was weak, so most of the students' emotions were positive and neutral, and a small portion of them were negative; in the diffusion period, due to the further expansion of the scope of influence of Topic2 and 3, which made the students' emotions peak at once, students for the event of anxiety, sadness, resentment increased dramatically, in this stage of the student's negative emotions rose to 67.58%, is the three stages of the total number of emotions and the proportion of the highest stage, the emotional value also fell to 0.1329; in the decline period, the student's emotion tends to decline, the proportion of the highest negative emotions also by the diffusion of the 67.58% fell to 14.28%, the positive emotion value also rose to 0.8124; at this time, due to the intervention of teachers and parents, the overall emotional characteristics of the students of different Civics themes situation has been clearly understood, the students' emotions began to gradually weaken, showing trends such as rationalization.

Table 5. Three stages of different emotion quantities and emotion values.

Emotion Evolution Stage	Positive emotions			Neutral emotions			Negative emotions		
	Emotion count	Percentage	Emotion value	Emotion count	Percentage	Emotion value	Emotion count	Percentage	Emotion value
Incubation	3574	28.90%	0.7827	1984	26.21%	0.5053	5683	18.14%	0.1874
Spread	7412	59.93%	0.7883	4658	61.54%	0.5019	21176	67.58%	0.1329
Winter	1382	11.17%	0.8124	927	12.25%	0.5011	4474	14.28%	0.1278
Tote	12368	100%	-	7569	100%	-	31333	100%	-

4.3.3. Thematic Emotional Analysis of Civic Education Contexts by Stage

In this section, based on the results after theme extraction, the sentiment score, i.e., the mean value of sentiment intensity, is calculated separately for each stage of the themes subordinate to the text comments of the same theme. The details are shown in Table 6.

As can be seen from the table, combined with the results after theme extraction, it can be seen that in the latency period, students for Topic6 (miracle, red, cheer, faith) express more emotional number of 5875, Topic4 (democracy, the West, taught, the motherland) has the highest value of 0.6317, combined with the extracted theme words, it can be seen that the students' attitudes at this time are more positive and optimistic, emotions more positive; in the diffusion period, students for Topic3 (Japan, humiliation, lead, self-improvement) of the number of emotions expressed more 12795, Topic2 (the United States, freedom, peace, strangulation) of the highest emotional value of 0.6409, can be seen that the students at this time a high degree of attention to the changes in the war of the countries; in the recession period, students for Topic7 (Shengshi, desired, martyrdom, Gratitude) has a higher sentiment number expression of 2576, which can be seen that students are highly concerned about the development of the country.

Table 6. Number and value of emotions in different themes in three stages.

Emotion Evolution Stage	Topic	Emotion count	Emotion value
Incubation	1	2147	0.2094
	2	1763	0.1985
	3	879	0.2273
	4	2473	0.6317
	5	957	0.1128
	6	5875	0.5162
	7	1184	0.1463
	8	1029	0.1546
Spread	1	2209	0.1872
	2	2664	0.6409
	3	12795	0.4731
	4	2972	0.1856
	5	3094	0.2358
	6	4781	0.2297
	7	1074	0.1058
	8	5039	0.2694
Winter	1	948	0.5732
	2	1483	0.1876
	3	2073	0.2148
	4	897	0.1082
	5	1139	0.1543
	6	1948	0.2054
	7	2946	0.4874
	8	2236	0.2312

5. Conclusion

In this paper, LDA topic model and feature-weighted plain Bayesian algorithm are used to extract the topics and analyze the emotional tendency of students' emotions in different ideological contexts, and the experimental results show that the proposed feature-weighted fusion of plain Bayesian emotion classification algorithm achieves an accuracy rate of 95.92% on binary data, and an accuracy rate of 82.76% on multicategory data, which is a better effect, with the performance ranked as Feature weighted plain Bayesian model > traditional machine learning model > sentiment dictionary classification method. Through word frequency statistics and the calculation and comparison of keyword TF-IDF values, the high-frequency words of the text are highlighted to deepen the understanding of the content of the text on Civic Education.

Based on the extracted thematic content of the LDA model, the thematic sentiment of the students' text data in each stage was calculated and analyzed using the feature-weighted fusion of the plain Bayesian sentiment classification model, in the latent period students' sentiment is more neutral and negative, in the diffusion period negative sentiment reaches the peak, however, with the increase of the concept of Civic and Political Education, in the decline period students' sentiment is becoming more and more peaceful, which shows that the design of this paper based on the feature-weighted fusion of the plain Bayesian emotion classification model designed in this paper effectively achieves the purpose of analyzing students' emotional tendencies.

About the Author

Xiaoying Liu (1985.1-), female, Han, Zhoushan, Zhejiang Province, is a lecturer at the College of Fashion & Design, Zhejiang Fashion Institute of Technology. Her research focuses on ideological and political education for students, student mental health, and career guidance. Xianyi Li (1980.1-), male, Han, from Guanxian, Shandong Province, Senior Researcher, School of Mechatronics and Energy Engineering, NingboTech University, His research interests include engineering practice teaching and intelligent manufacturing.

References

1. Dong, F., & Dong, S. (2023). Research on the optimization of ideological and political education in universities integrating artificial intelligence technology under the guidance of curriculum ideological and political thinking. ACM Transactions on Asian and Low-Resource Language Information Processing.
2. Du, G., Sun, Y., & Zhao, Y. (2023). The innovation of ideological and political education integrating artificial intelligence big data with the support of wireless network. Applied artificial intelligence, 37(1), 2219943.
3. Liu, Z., & Luo, L. (2024, March). Using Artificial Intelligence for Intelligent Ideological and Political Education Teaching. In 2024 International Conference on Interactive Intelligent Systems and Techniques (IIST) (pp. 137-144). IEEE.
4. Qing, Z. (2024). Innovating Content and Methods of Ideological and Political Education in the Context of the New Era. International Journal of Education and Humanities, 4(4), 479-487.
5. Zhang, T., Lu, X., Zhu, X., & Zhang, J. (2023). The contributions of AI in the development of ideological and political perspectives in education. Heliyon, 9(3).
6. Liu, Z., & Huang, Y. (2024). Innovation in Ideological and Political Education and Personalized Learning Paths in the Era of Artificial Intelligence. International Journal of New Developments in Education, 6(5), 178-183.
7. Xiaoyang, H., Junzhi, Z., Jingyuan, F., & Xiuxia, Z. (2021). Effectiveness of ideological and political education reform in universities based on data mining artificial intelligence technology. Journal of Intelligent & Fuzzy Systems, 40(2), 3743-3754.
8. Wang, L., Hu, G., & Zhou, T. (2018). Semantic analysis of learners' emotional tendencies on online MOOC education. Sustainability, 10(6), 1921.
9. Deng, J. (2021, January). Analysis and Countermeasures of College Students' Sentimental Tendency Based on Network Behavior Data. In 2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) (pp. 653-656). IEEE.
10. Moreno-Ger, P., & Verdú, E. (2023). A Ubiquitous Model of Emotional Tracking in Virtual Classes: From Simple Emotions to Learning Action Tendency. IEEE Latin America Transactions, 21(8), 889-896.
11. Zhang, J., Zhou, Y., Gong, Y., & Tang, X. (2024, March). Analysis of Student Classroom Emotional Tendency Based on Random Forest Model. In 2024 13th International Conference on Educational and Information Technology (ICEIT) (pp. 240-246). IEEE.
12. Yuzhong, H. (2021). Students' emotional analysis on ideological and political teaching classes based on artificial intelligence and data mining. Journal of Intelligent & Fuzzy Systems, 40(2), 3801-3809.
13. Qiujiang, Y. (2024, August). Research on Feedback System of Ideological and Political Teaching Resource Base Based on Emotion Analysis and Deep Learning. In 2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE) (pp. 1730-1734). IEEE.
14. Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. Soft Computing, 25(3), 2277-2293.

15. Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics, 403(412).
16. Gao, H., Zeng, X., & Yao, C. (2019). Application of improved distributed naive Bayesian algorithms in text classification. Journal of Supercomputing, 75(9).
17. Abbas, M., Memon, K. A., Jamali, A. A., Memon, S., & Ahmed, A. (2019). Multinomial Naive Bayes classification model for sentiment analysis. IJCSNS Int. J. Comput. Sci. Netw. Secur, 19(3), 62.
18. Li, Z., Li, R., & Jin, G. (2020). Sentiment analysis of danmaku videos based on naïve bayes and sentiment dictionary. Ieee Access, 8, 75073-75084.
19. Yang Hulin & Zhu Ai'nan. (2021). Analysis of Telecommunication Fraud Cases Based on TF-IDF Algorithm.IOP Conference Series: Earth and Environmental Science,742(1),https://doi.org/10.1088/1755-1315/742/1/012011.
20. Mingfeng Jiang. (2025). Semantic Data Mining for Reading Promotion News Based on LDA Topic Models.Forum on Research and Innovation Management,3(9),https://doi.org/10.70711/FRIM.V3I9.7332.
21. Khanh N. Dinh,Cécile Liu,Zijin Xiang,Zhihan Liu & Simon Tavaré. (2025). Approximate Bayesian computation sequential Monte Carlo via random forests.Statistics and Computing,35(6),219-219.https://doi.org/10.1007/S11222-025-10748-X.