Hindawi
Discrete Dynamics in Nature and Society
Volume 2024, Article ID 9895382, 1 page
https://doi.org/10.1155/2024/9895382

*Retraction*

# Retracted: Semantic-Based Classification of Long Texts on Higher Education in China

**Discrete Dynamics in Nature and Society**

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024



This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

## References

[1] C. Li and Y. Fei, "Semantic-Based Classification of Long Texts on Higher Education in China," *Discrete Dynamics in Nature and Society*, vol. 2021, Article ID 9237713, 8 pages, 2021.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2021, Article ID 9237713, 8 pages
https://doi.org/10.1155/2021/9237713

*Research Article*

# Semantic-Based Classification of Long Texts on Higher Education in China

**Chun Li [1] and Yanying Fei [2]**

[1]School of Marxism, Dalian University of Technology, Dalian 116023, China
[2]Faculty of Humanities and Social Sciences, Dalian University of Technology, Dalian 116086, China

Correspondence should be addressed to Chun Li; lidoctor@mail.dlut.edu.cn

Received 14 September 2021; Revised 11 October 2021; Accepted 12 October 2021; Published 20 October 2021

Academic Editor: Gengxin Sun



The development level of higher education (HE) is an important indicator of the development level and development potential of a country. The HE-related document is the mirror to reflect the develop process of the HE. The research of high education (HE) has been developing rapidly in China, resulting in a huge number of texts, such as relevant policies, speech drafts, and yearbooks. The traditional manual classification of HE texts is inefficient and unable to deal with the huge number of HE texts. Besides, the effect of direct classification is rather poor because HE texts tend to be long and exist as an imbalanced dataset. To solve these problems, this paper improves the convolutional neural network (CNN) into the HE-CNN classification model for HE texts. Firstly, Chinese HE policies, speech drafts, and yearbooks (1979–2020) were downloaded from the official website of Chinese Ministry of Education. In total, 463 files were collected and divided into four classes, namely, definition, task, method, and effect evaluation. To handle the huge number of HE texts, the Twitter-latent Dirichlet allocation (LDA) topic model was employed to extract word frequency and critical information, such as age and author, enhancing the training effect of CNN. To address the dataset imbalance problem, CNN parameters were optimized repeatedly through comparative experiments, which further improve the training effect. Finally, the proposed HE-CNN model was found more effective and accurate than other classification models.



## 1. Introduction

Higher education (HE) is a crucial part of the country education system [1] and the foundation of national talent training. In recent years, with the continuous development of China's national power, higher education has also developed rapidly and the HE serves the national development. Therefore, through semantic analysis of higher education-related documents, the development trend of higher education can be discovered, and future planning and research can be carried out.

The HE has been developing rapidly in China, resulting in a huge number of texts, such as relevant policies, speech drafts, and yearbooks. The traditional research almost uses the manual method (some statistic methods) in analysis of higher education-related data in the field of humanities and social sciences. However, with the development of HE research, the semantic-based HE document analysis model should be used in future HE research aspect to solve the problem of HE texts, such as inefficient and hard to process. What is worse, there is no classification model for HE files data in any academic database. In order to solve this problem, this paper intends to develop a semantic-based classification model for HE texts.

In 2020s, the social networking, which uses the semantic-based text analysis, becomes a hot topic in the field of computer science [2]. Various accurate text mining models have emerged, including convolutional neural network (CNN) and long short-term memory (LSTM) model [3–12]. However, these traditional classification models cannot be directly used for training and applying directly to process the HE files because the HE files are much longer and richer in contents, and the whole HE dataset is more imbalanced than common social network texts (e.g., Tweets) [13].

To address the above problems, this paper firstly sets up a standard Chinese HE dataset, using a Python crawler. The dataset includes the policies, yearbooks, and speed drafts on HE in 1979–2020. In total, there are 466 files in the dataset. Every text file is long text style, which contain thousands of words at least. So, although the number of text file is not too large, it is still hard to build the text analysis model. The code and data will be uploaded to GitHub in future. Next, the CNN was extended into an HE classification model called HE-CNN. (1) The huge number of HE files makes it hard to train the classification model. To solve this problem, the Twitter-latent Dirichlet allocation (LDA) topic model was adopted to extract and compress text data, convert long texts into short texts, and then compress the texts, without sacrificing the critical contents. (2) To solve the dataset imbalance problem, the CNN training effect was improved with a mixture of texts and special attributes (e.g., period and high-frequency words). In addition, the parameters of the HE-CNN model were optimized experimentally through cross validation, making the classification more accurate and training more efficient. Experimental results show that the optimized model strikes a good balance between accuracy and efficiency, compared with unoptimized classification models. The main contributions of this research are as follows:

(1) The CNN classification model was improved into the HE-CNN classification for HE files. The proposed model handles the long texts in HE files through keyword extraction and overcomes the dataset imbalance problem by expanding the training set with text attributes. Moreover, the model parameters were tuned to balance training time with classification accuracy, thereby improving the model training effect. This is unachievable with traditional CNN.

(2) A standard Chinese HE dataset of 466 files, including speech drafts, policies, and yearbooks, was established, laying a solid data basis for future attempts of HE classification.

## 2. Literature Review

For the semantic analysis problem, there are some works which have achieved great successes results in processing the short social network text (Twitter and Weibo). For instance, Yue et al. [14] designed a classification model for short texts like tweets and tax invoices. Relying on Chinese knowledge graph, their model solves the sparsity of data labels and facilitates model training. Gultepe et al. [15] presented a CNN-based simple classifier for text files: a new CNN architecture is adopted to utilize locally trained latent semantic analysis (LSA) word vectors. Qiu et al. [16] proposed a multichannel semantic synthesis CNN (SFCNN). To complete the task of emotion classification, the emotional weights of word vectors are determined through multichannel semantic synthesis, and the model parameters are optimized through gradient descent with an adaptive learning rate. Shi and Zhao [17] developed a semantic classifier based on a neural network. The theoretical findings of the axiom fuzzy set theory are incorporated to the neural network, and complex concepts are extracted by the neural network to enhance classification accuracy. Wang et al. [18] put forward a dual feature training support vector machine (SVM) model to classify texts and images. Wu et al. [19] proposed a deep semantic matching model, which finetunes CNN parameters through the generation of candidate entities. Yang et al. [20] came up with two semantic-based Chinese file classification strategies: the ambiguity problem is solved by a novel semantic similarity calculation (SSC) method and the problem of synonyms is overcome through a robust correlation analysis method (SCM).

Apart from semantic-based text classification, the text mining model becomes a new hotspot in the field of education. For instance, Chen et al. [21] classified questionnaire data with a semantic analysis model to solve the educational backwardness of ethnic minorities. Goncalves et al. [22] used a semantic model to evaluate massive open online courses (MOOC) and improved the course quality against the classification results. Niu et al. [23] proposed a novel theory of semantic cohesion for Chinese airworthiness regulations and specified four critical elements of the theory, namely, definition, model, theorem, and rules. Koutsomitropoulos et al. [24] combined explicit knowledge graph representations with vector-based learning of formal thesaurus terms into a hybrid semantic classification model and demonstrated the good effect of the hybrid model on the classification of biological files in library terminology learning. With the aid of the enhanced learning model, Shen and Ho [25] evaluated HE teaching effects and quickly grasped the development state of HE.

In summary, semantic-based text analysis is relatively mature in computer science but remains in the early stage in the field of HE. Therefore, this paper aims to improve semantic-based analysis on HE text classification.

## 3. Data Collection and Preprocessing

Our data fall into three categories: yearbooks, speech drafts, and policies. The original data were mainly downloaded from the open datasets provided on the official website of Chinese Ministry of Education (https://www.moe.gov.cn/jyb_sjzl/moe_364/zgjynj_2015/), using a self-designed Python crawler. The HyperText Markup Language (HTML) tags were removed to generate the final experimental dataset. In total, the dataset contains 466 HE-related files of speech drafts and policies and 331 HE yearbook files, all of which were released between 1988 and 2019 (Figures 1 and 2).

There are two primary attributes in the data: period and high-frequency words. The period words are the label that can reflect the key words changed with the time, and high-frequency words can partially summarize the main topic of the document.

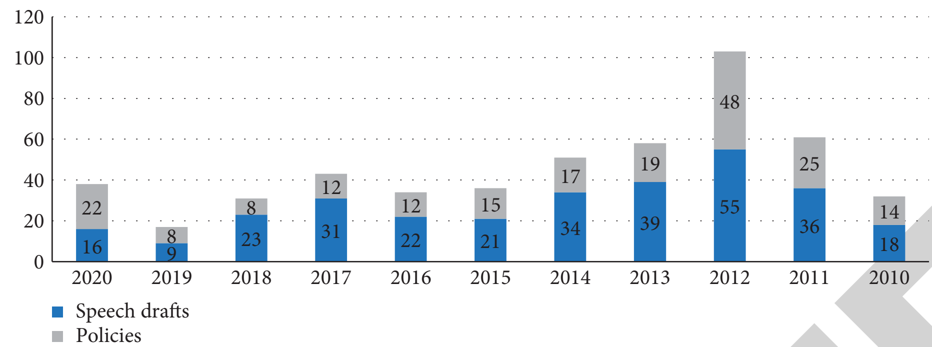


Figure 1: Data on HE speech drafts and policies.

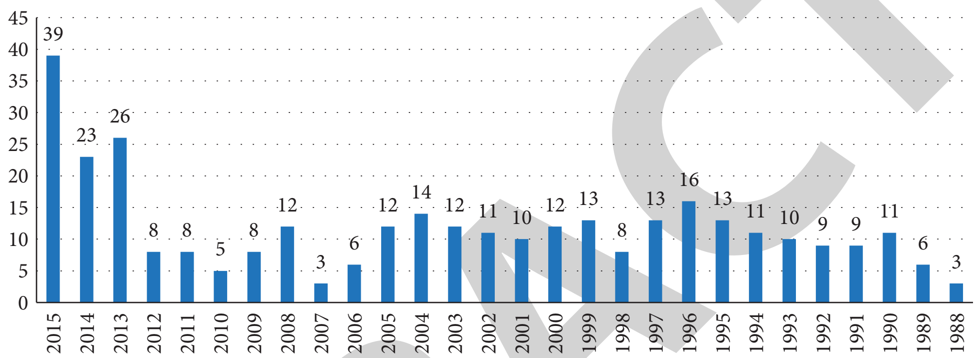

Figure 2: Data on HE yearbooks.

Table 1: The entire experimental dataset.

| Description | | Definition | Task | Method | Effect evaluation |
|---|---|---|---|---|---|
| Comprehensive experiment | Training set | 160 | 156 | 151 | 171 |
| | Validation set | 40 | 44 | 35 | 43 |
| | Test set | 40 | 44 | 46 | 30 |
| File feature extraction | | 797 texts | | | |

The two attributes were adopted to enhance the classification accuracy of our semantic-based model. The entire experimental dataset is illustrated in Table 1.

## 4. Text Attributes and Model Framework

This section introduces the overall framework of the proposed HE-CNN model. As its name suggests, our model consists of two parts: one is the traditional CNN and the other is the attributes of HE files (i.e., year and high-frequency words). The framework of the proposed model is shown in Figure 3.

*4.1. Text Mining Model.* In 2014, Kim [26] proposed the CNN model for text classification. Their model converts the original text into multidimensional vectors for further analysis and achieves relatively good classification results. However, the model requires a huge amount of training data and does poorly in fault tolerance. In recent years, several novel methods have been introduced to optimize the CNN model. For example, Zhang et al. [27] proposed a three-way model that improves the classification accuracy of CNN for emotional texts. Yang et al. [28] employed a multichannel SFCNN to overcome the emotional ambiguity caused by the changing text context.

In this paper, 100 convolutional filters (size: 1) are adopted to process text vectors, each text is segmented into words by Python Jieba, and max pooling is performed to analyze the output, which is cascaded from the results of all filters. The text representation model is shown in Figure 4.

*4.2. File Feature Extraction.* This paper relies on the unique features of each HE file (i.e., publish year and main topic) to improve HE-CNN training performance and overcome dataset imbalance.

The publish year was selected for two reasons. (1) The HE development in China is greatly affected by government policy. The HE files usually reflect the concept of governance. In the 1990s, HE mainly emphasized on vocational education, a booster of industrialization. In the 2000s, the

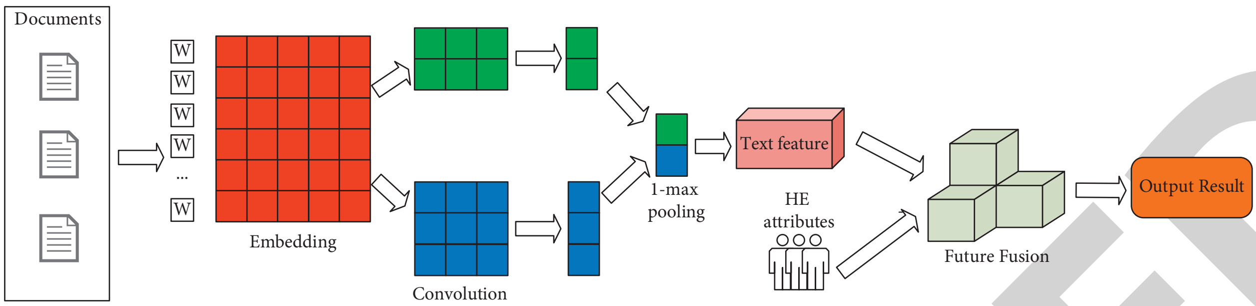


Figure 3: Framework of HE-CNN.

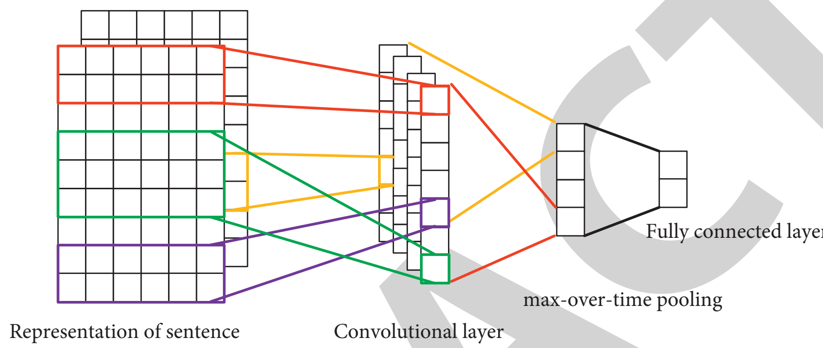


Figure 4: Text representation model.



focus of HE gradually shifted towards science and technology. The shift is a response to China's requirements on HE in different periods. (2) The publish year is readily available in the files. Here, the attribute of publish year is divided by decade into the 1980s, the 1990s, the 2000s, and the 2010s.

The file features cannot be directly extracted from the files. Thus, the semisupervised Twitter-LDA topic model [29] was selected to extract the topic from each file. In this paper, in order to highly generalize the document, the first five high-frequency keywords are chosen as the topic words to describe the file. Once extracted, the eigenvector of each file was taken as the topic vector, and the topic vectors of all files were merged and divided by publishing year to facilitate model training.

Finally, a soft layer was added to HE-CNN to output the result. All layers of our model were processed by a normalization algorithm, such that the parameters between different layers can be dynamically adjusted with the training data. The model parameters are listed in Table 2.

## 5. Experimental Results

This section mainly reports the experimental results on our model. Firstly, the experimental results and effects are measured by the low function as follows:

$$L = -\sum_{i=1}^{N} y_i \log(p_i), \tag{1}$$

Table 2: Model parameters.

| Description | Value |
|---|---|
| Word representation | Static 300-dimensional word2vec |
| Size of convolution kernel | 1 |
| Number of convolution kernels | 100 |
| Pooling | 1-max pooling |
| Dropout | 0.5 |

where $N$ is the number of semantic classes; $p_i$ is the value of the i-th output vector; and $y_i$ is the ground truth.

*5.1. Parameter Configuration.* Before any experiment, HE-CNN parameters must be initialized and optimized, for classification accuracy hinges on feature representation. In this paper, the model parameters are optimized by the word2vec model. The input layer of our model was trained separately in multiple modes, namely, 100-dimensional, 200-dimensional, 300-dimensional, and 400-dimensional vectors, using the text data on yearbooks, speech drafts, and policies. The training results indicate that the 300-dimensional experiment had the best effect. Hence, the 300-dimensional vector was taken as the parameter of the input layer.

For the convolution layer, 100 kernels were selected after multiple experiments. Once the number of kernels surpassed

Table 3: Classification performance at different numbers of kernels.

| Number of kernels | Accuracy (%) | Time cost (s) |
|---|---|---|
| 10 | 69.70 | 1945 |
| 20 | 72.46 | 2087 |
| 40 | 75.55 | 2158 |
| 60 | 78.78 | 2274 |
| 80 | 80.28 | 2399 |
| 100 | 82.15 | 2458 |
| 120 | 82.22 | 2732 |
| 140 | 82.46 | 3058 |

100 and continued to grow, the classification accuracy did not increase, but the training time surged up (Table 3).

For the pooling layer, 1-max pooling achieved the best performance. The dropout rate had a small effect on the model and was thus set to 0.5 for our experiments. Table 2 lists the model parameters for experiments. In addition, the LSTM model was also adopted to extract text features in our experiments.

*5.2. Attribute Extraction.* The feature distribution of each class was extracted from all 463 files in four periods. Firstly, the Twitter-LDA topic model was employed to extract the distribution of class preferences. Focusing on all the files published in a period, the model holds that the features of each class are reflected by high-frequency words. After adjusting the number of topics $k$ to 6, each of the six topics was labeled (i.e., education, development, construction, reform, school, and party). Under the six topics, 30 keywords (translated literally into English) were sorted by frequency. The results in Table 4 suggest that these keywords are reasonably grouped under corresponding topics.

To make each topic more intuitive and facilitate the analysis of period distribution, the topic word distribution was fixed, regardless of periods, to analyze preference distribution. As shown in Figure 5, the preference distribution across periods varied greatly from topic to topic. For example, Figure 6 displays the preference distribution on various topics in the 2010s. The preference distribution was adopted as the vectorized representation of a file feature and combined with publish year (four dimensions: 1980s, 1990s, 2000s, and 2010s) and high-frequency words (five dimensions) of the files. In this way, a 20-dimensional feature was obtained to depict file attributes.

*5.3. Comparative Experiments.* Random cross validation was implemented in model training. Specifically, 25% of the training data were used as the cross-validation set, and 40% standard files were used as the validation set for each validation. The fusion model was trained for 200 generations, and each trained model was validated for 204 iterations. After a total of 40,800 iterations, the stability of the fusion model was evaluated against the test set by

$$\mathrm{Acc} = \frac{\sum_{i=1}^{N} (\mathrm{TP} + \mathrm{TN}/\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})}{N}, \qquad (2)$$

where $N$ is the number of semantic classes; TP is the true positive; TN is the true negative; FP is the false positive; and FN is the false negative.

For comparison, our model was contrasted with CNN, decision tree (DT), Naïve Bayes (NB), k-nearest neighbors (KNN), random forest (RF), multilayer perceptron (MLP), SVM, and logistics regression (LR). The results (Table 5) show that our model far outperformed these traditional classification models on the same HE dataset.

As mentioned before, HE files are generally composed of a few long texts, each of which is very large. To ensure the sufficiency of training, standard text dataset and text features were combined in the training set. The performance of different classification models on the HE dataset with text features is compared in Table 6. It can be seen that for the long text style document, the CNN model obtains a better result than other models. In addition, using the extracted text features can significantly improved the performance of the HE-CNN model than traditional CNN models.

Furthermore, many combinations of text-feature presentation dimensions were tested to optimize the CNN parameters. In order to keep the balance between the training speed and model accuracy, the optimal experiment is processed and the result is shown in Table 7. The results in Table 7 suggest that the number of the text-feature dimension is 300 which can get the balance.

In addition, the experiment also uses the text-feature LSTM model to extract the text features which could increase the number of the text features. However, the experiment reflected that the extracted futures using text-feature LSTM model cannot be significantly used for improving the accuracy and indicated that the LSTM model is not suitable for processing long texts. The final experimental results are given in Table 8. Apparently, our HE-CNN is applicable to semantic-based HE text analysis and is fast in dividing the texts into different classes.

Table 4: Keyword distribution of each topic.

| Topic | Results | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Education | Country | 0.00008 | Important | 0.00315 | Training | 0.00979 | Employment | 0.00475 | Focus | 0.00507 |
| | Society | 0.0013 | Question | 0.0044 | Economy | 0.0036 | Management | 0.0057 | National | 0.0074 |
| | Student | 0.0019 | Research | 0.0098 | Science | 0.0012 | Quality | 0.007 | Need | 0.0046 |
| | Innovation | 0.005 | Teacher | 0.0094 | Thought | 0.0034 | Morality | 0.0067 | Government | 0.0041 |
| | Our country | 0.0097 | Rural | 0.0035 | China | 0.0058 | Culture | 0.0077 | Mechanism | 0.0043 |
| | Improvement | 0.0033 | Talent | 0.0016 | System | 0.0052 | Services | 0.0092 | Level | 0.0097 |
| Development | Education | 0.0092 | Improve | 0.00248 | Economic | 0.00875 | Implement | 0.0043 | Focus | 0.00462 |
| | Construction | 0.00103 | Cause | 0.0005 | Socialism | 0.00778 | Talent | 0.00313 | Quality | 0.0056 |
| | Reform | 0.00665 | Problem | 0.00225 | Innovation | 0.00786 | Level | 0.00105 | Service | 0.00393 |
| | Society | 0.00905 | China | 0.00509 | Science | 0.00759 | Need | 0.00979 | Student | 0.0094 |
| | Country | 0.00207 | School | 0.00584 | Bring up | 0.00378 | The study | 0.00544 | System | 0.00425 |
| | Our country | 0.00222 | Promote | 0.00988 | Colleges | 0.00703 | Features | 0.00787 | People | 0.00238 |
| Construction | Strategy | 0.00131 | Must | 0.0053 | Advance | 0.00721 | Postgraduate | 0.00128 | Modernization | 0.00528 |
| | Cooperation | 0.00693 | Science and technology | 0.00297 | Comprehensive | 0.00613 | Communication | 0.00174 | Our country | 0.00724 |
| | Strengthen | 0.00388 | Cadre | 0.00871 | Xi Jinping | 0.00405 | Philosophy | 0.00396 | Innovation | 0.00489 |
| | Organization | 0.00868 | Intellectuals | 0.00023 | System | 0.00395 | Social science | 0.00688 | Reform | 0.00095 |
| | Nation | 0.00332 | Produce | 0.00461 | Accelerate | 0.00468 | Theory | 0.00054 | Development | 0.00593 |
| | Nationwide | 0.00231 | System reform | 0.00795 | General secretary | 0.00868 | Socialism | 0.00531 | Thought | 0.00652 |
| Reform | Nationwide | 0.00339 | Society | 0.00578 | Thought | 0.00154 | People | 0.00273 | Colleges | 0.00335 |
| | Education | 0.002 | Modernization | 0.00696 | Practice | 0.00969 | Service | 0.00453 | Profession | 0.00165 |
| | China | 0.00102 | Construction | 0.00083 | Problem | 0.00794 | Political | 0.00394 | System | 0.00453 |
| | Features | 0.00597 | Our country | 0.00638 | Science | 0.00354 | Central | 0.00588 | School | 0.00419 |
| | Socialism | 0.00719 | Innovation | 0.00819 | Leadership | 0.00215 | Solve | 0.00775 | All levels | 0.009 |
| | Cause | 0.00865 | Development of | 0.00546 | School | 0.00532 | Management | 0.00889 | The study | 0.00325 |
| School | Bring up | 0.0097 | Thought | 0.00096 | Our country | 0.00225 | Modernization | 0.00785 | Improve | 0.00428 |
| | Cause | 0.00642 | People | 0.00533 | Leadership | 0.00675 | School | 0.00135 | Central | 0.00076 |
| | China | 0.00047 | Development of | 0.00988 | Country | 0.00112 | Implement | 0.00184 | Political | 0.0027 |
| | Culture | 0.00694 | Socialism | 0.00953 | System | 0.00851 | Government | 0.00617 | All levels | 0.00372 |
| | Practice | 0.0029 | Education | 0.00374 | Features | 0.00566 | Society | 0.00686 | Reform | 0.00363 |
| | Promote | 0.0015 | Problem | 0.00264 | Talent | 0.00098 | Student | 0.00866 | The study | 0.00068 |
| Party | Important | 0.00625 | Cause | 0.00326 | Leadership | 0.00747 | Solve | 0.00217 | Practice | 0.00675 |
| | Employment | 0.00312 | Implement | 0.00959 | Promote | 0.00786 | Teaching | 0.00399 | Enterprise | 0.00369 |
| | Mechanism | 0.00897 | School | 0.00547 | Technology | 0.00651 | Technology | 0.0028 | International | 0.00578 |
| | Colleges | 0.00245 | Political | 0.00667 | Plan | 0.0003 | Central | 0.00668 | Major | 0.00833 |
| | Profession | 0.00418 | People | 0.00536 | Modernization | 0.00724 | Department | 0.00222 | World | 0.00601 |
| | Socialism | 0.00844 | Area | 0.00169 | All levels | 0.00835 | Features | 0.00498 | Engineering | 0.00913 |

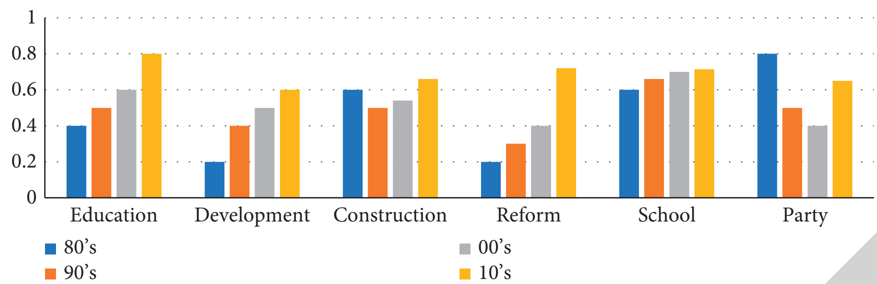


Figure 5: Preference distribution of the six topics across periods.

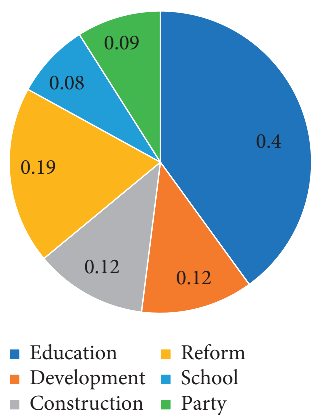


Figure 6: Preference distribution in the 2010s.

Table 5: Classification performance of different models on our HE dataset.

| Model | Accuracy (%) |
|---|---|
| DT | 72.746 |
| NB | 73.856 |
| KNN | 60.555 |
| RF | 52.568 |
| MLP | 34.583 |
| SVM | 61.734 |
| LR | 72.786 |
| CNN | 78.56 |
| HE-CNN | 83.516 |

Table 6: Classification performance of different models on our HE dataset with text features.

| Model | Accuracy (%) |
|---|---|
| DT | 77.424 |
| NB | 75.824 |
| KNN | 64.236 |
| RF | 54.358 |
| MLP | 33.528 |
| SVM | 68.157 |
| LR | 75.856 |
| CNN | 81.356 |
| HE-CNN | 88.782 |

Table 7: Accuracy of different presentation dimensions of the text features.

| Model | Accuracy (%) |
|---|---|
| Text-100 | 46.213 |
| Text 100-fusion | 79.745 |
| Text 200 | 69.532 |
| Text 200-fusion | 79.212 |
| Text 300 | 76.332 |
| Text 300-fusion | 82.345 |
| Text 400 | 69.712 |
| Text 400-fusion | 78.662 |

Table 8: Final experimental results.

| Model | Accuracy (%) |
|---|---|
| Text CNN | 73.323 |
| Text LSTM | 66.755 |
| Feature CNN | 31.875 |
| Text-feature CNN | 77.587 |
| Text-feature LSTM | 67.976 |
| Our model | 88.782 |

## 6. Conclusions

To solve the classification problem of HE texts, this paper builds a standard HE dataset and proposes the HE-CNN model, which combines text features with optimal CNN parameters. The proposed dataset lays the foundation for future studies on HE classification models, while our model was designed to handle large HE datasets containing long texts. The proposed model was proved effective through comprehensive experiments.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This research was supported by "Social risks in the development of artificial intelligence industry in Liaoning and their legal countermeasures" (LNFXH2020A017), Key Project of Liaoning Provincial Law Society, 2020-2021.

## References

[1] D. H. Chung, M. Jin, J. Mina, and P. Kyeong-Jin, "Analysis of the questioning characteristics of elementary science gifted education teaching materials using the sternberg's view of successful intelligence: focused on semantic network analysis," *Journal of the Korean Earth Science Society*, vol. 40, no. 6, pp. 654–670, 2019.

[2] B. Jia, B. Meng, W. Zhang, and J. Liu, "Query rewriting and semantic annotation in semantic-based image retrieval under heterogeneous ontologies of big data," *Traitement du Signal*, vol. 37, no. 1, pp. 101–105, 2020.

[3] P. K. Yechuri and S. Ramadass, "Semantic web mining for analyzing retail environment using Word2Vec and CNN-FK," *Ingénierie des Systèmes d'Information*, vol. 26, no. 3, pp. 311–318, 2021.

[4] S. Luo, Y. Gu, X. Yao, and W. Fan, "Research on text sentiment analysis based on neural network and ensemble learning," *Revue d'Intelligence Artificielle*, vol. 35, no. 1, pp. 63–70, 2021.

[5] W. Lejmi, A. B. Khalifa, and M. A. Mahjoub, "A novel spatio-temporal violence classification framework based on material derivative and LSTM neural network," *Traitement du Signal*, vol. 37, no. 5, pp. 687–701, 2020.

[6] C. Zhang, Q. Li, and X. Cheng, "Text sentiment classification based on feature fusion," *Revue d'Intelligence Artificielle*, vol. 34, no. 4, pp. 515–520, 2020.

[7] J. Xie, R. Li, S. Lv, Y. Wang, Q. Wang, and Y. Vorotnitsky, "Chinese alt text writing based on deep learning," *Traitement du Signal*, vol. 36, no. 2, pp. 161–170, 2019.

[8] B. Ilyes, V. Frederic, H. Denis, and D. Fadi, "Multi-label, multi-task CNN approach for context-based emotion recognition," *Information Fusion*, vol. 76, pp. 422–428, 2021.

[9] S. Shengli, H. Haitao, and R. Tongxiao, "Abstractive text summarization using LSTM-CNN based deep learning," *Multimedia Tools and Applications*, vol. 78, no. 1, pp. 857–875, 2019.

[10] Z. Yangsen, Z. Jia, J. Yuru, H. Gaijuan, and C. Ruoyu, "A text sentiment classification modeling method based on coordinated CNN-LSTM-Attention model," *Chinese Journal of Electronics*, vol. 28, no. 1, pp. 120–126, 2019.

[11] K. Tauseef and M. Ayatullah Faruk, "AUTNT - a component level dataset for text non-text classification and benchmarking with novel script invariant feature descriptors and D-CNN," *Multimedia Tools and Applications*, vol. 78, no. 22, pp. 32159–32186, 2019.

[12] B. Imon, L. Yuan, C. Matthew C, and H. Sadid, "Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification," *Artificial Intelligence in Medicine*, vol. 97, pp. 79–88, 2019.

[13] H. Casakin and V. Singh, "Insights from a latent semantic analysis of patterns in design expertise: implications for education," *Education Sciences*, vol. 9, no. 3, p. 23, 2019.

[14] Y. Yue, Y. Zhang, X. Hu, and P. Li, "Extremely short Chinese text classification method based on bidirectional semantic extension," *Journal of Physics: Conference Series*, vol. 1437, no. 1, Article ID 012026, 2020.

[15] E. Gultepe, M. Kamkarhaghighi, and M. Makrehchi, "Document classification using convolutional neural networks with small window sizes and latent semantic analysis," *Web Intelligence*, vol. 18, no. 3, pp. 239–248, 2020.

[16] N. Qiu, S. Zhou, L. Cong, P. Wang, and Y. Li, "Research on multi-channel semantic fusion emotion classification model based on CNN," *Computer Engineering and Applications*, vol. 55, no. 23, pp. 136–141, 2019.

[17] Y. Shi and J. Zhao, "The semantic classification approach base on neural networks," *IEEE Access*, vol. 8, pp. 14573–14578, 2020.

[18] Y. Wang, W. Yu, and Z. Fang, "Multiple kernel-based SVM classification of hyperspectral images by combining spectral, spatial, and semantic information," *Remote Sensing*, vol. 12, no. 1, p. 120, 2020.

[19] X. Wu, Y. Duan, Y. Zhang, and X. Yan, "A Chinese entity linking model based on CNN and deep structured semantic model," *Computer Engineering and Science*, vol. 42, no. 8, pp. 1514–1520, 2020.

[20] S. Yang, R. Wei, J. Guo, and H. Tan, "Chinese semantic document classification based on strategies of semantic similarity computation and correlation analysis," *Journal of Web Semantics*, vol. 63, p. 100578, 2020.

[21] X. Chen, C. Tian, and T. Wu, "The semantic web approach for the collaborative analysis and visualization of ethnic education and vocation," in *Proceedings of the 2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pp. 414–419, Nanjing, China, May 2018.

[22] V. Gonçalves, B. Gonçalves, and F. Garcia-Tartera, "MOOCs to semantic web education," in *Proceedings of the IX International Conference The Future of Education*, vol. 9, pp. 191–196, Florence, Italy, June 2019.

[23] H. Niu, C. Ma, P. Han, S. Li, and Q. Ma, "A novel semantic cohesion approach for Chinese airworthiness regulations: theory and application," *IEEE Access*, vol. 8, pp. 227729–227750, 2020.

[24] D. A. Koutsomitropoulos, A. D. Andriopoulos, and S. D. Likothanassis, "Semantic classification and indexing of open educational resources with word embeddings and ontologies," *Cybernetics and Information Technologies*, vol. 20, no. 5, pp. 95–116, 2020.

[25] C.-w. Shen and J.-t. Ho, "Technology-enhanced learning in higher education: a bibliometric analysis with latent semantic approach," *Computers in Human Behavior*, vol. 104, p. 106177, 2020.

[26] Y. Kim, "Convolutional neural networks for sentence classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1746–1751, Doha, Qatar, October 2014.

[27] Y. Zhang, Z. Zhang, D. Miao, and J. Wang, "Three-way enhanced convolutional neural networks for sentence-level sentiment classification," *Information Sciences*, vol. 477, pp. 55–64, 2019.

[28] D. Yang, N. Qiu, N. Qiu, L. Cong, and H. Yang, "Research on multi-channel semantic fusion classification model," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 23, no. 6, pp. 1044–1051, 2019.

[29] W. X. Zhao, J. Jiang, J. Weng et al., "Comparing twitter and traditional media using topic models," *Lecture Notes in Computer Science*, pp. 338–349, 2011.