Hindawi
Journal of Function Spaces
Volume 2023, Article ID 9870920, 1 page
https://doi.org/10.1155/2023/9870920

Retraction

Retracted: Construction of a Knowledge Map Based on Text CNN Algorithm for Maritime English Subjects

Journal of Function Spaces

Received 15 August 2023; Accepted 15 August 2023; Published 16 August 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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) Peer-review manipulation

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] H. Wang and A. Wei, "Construction of a Knowledge Map Based on Text CNN Algorithm for Maritime English Subjects," *Journal of Function Spaces*, vol. 2022, Article ID 6578682, 9 pages, 2022.

Hindawi
Journal of Function Spaces
Volume 2022, Article ID 6578682, 9 pages
https://doi.org/10.1155/2022/6578682

Research Article

Construction of a Knowledge Map Based on Text CNN Algorithm for Maritime English Subjects

Hui Wang and Aimin Wei

Marine School, Hainan Vocational University of Science and Technology, Haikou, Hainan 571126, China

Correspondence should be addressed to Aimin Wei; captalf@hvust.edu.cn

Received 16 May 2022; Revised 13 July 2022; Accepted 15 July 2022; Published 31 July 2022

Academic Editor: Miaochao Chen

Knowledge map is a new method of knowledge management with the information revolution. This paper is aimed at forming a systematic and standardized huge redundant knowledge structure, which can be used to mine the knowledge structure and the relationship between knowledge and visualize it in a graphical way, in order to obtain more representative information and improve the classification accuracy of text classification model. In this paper, a knowledge map construction method based on the Text CNN algorithm is proposed for the subject of Nautical English. It is of practical significance and academic value to make use of knowledge map to study Chinese Maritime English, which is helpful to the development of Chinese Maritime English and provides guidance. In order to maintain the diversity of particle swarm optimization, the Text CNN algorithm is combined with the construction of Maritime English subject knowledge map, and the network parameters and structure are optimized. Using knowledge map to study China Maritime English has important practical significance and academic value and has certain guiding significance for the development of China Maritime English.

1. Introduction

Many foreign ship owners prefer to spend more money to hire Indian or Filipino crew with better English ability but less business ability [1]. As an important part of China's transportation industry, the maritime industry has a unique spatial and temporal scope of activities, is at the forefront of reform and opening up, and is characterized by participation in international logistics, international conventions and practices, internationalism with a profound world background, and extensive international commonality [2]. In the face of the increasingly fierce competition in the international crew labor market, the level of English proficiency of crew members has become an important factor to improve the competitiveness of Chinese crew members, which is directly related to the internationalization of China's maritime industry in the 21st century [3]. Accordingly, the teaching and training of listening and speaking English for maritime professionals require students to be able to use the target language to complete practical work tasks in a realistic working environment with the help of the knowledge map of the maritime English subject and to improve the practical language skills [4].

The development of the discipline of Nautical English has now entered a mature stage; especially with the integration of postmodernist ideas, environmental-ecological perspectives, and ethical ideas, the content is constantly enriched, the number of literature increases all the time, and what is changing at the same time is the structure of the discipline's knowledge; the distribution of the author's academic community is gradually complicated and diversified with time and geography [5]. In addition, most students do not have a good memorization or learning method and just learn and memorize English knowledge points in a stereotypical and mechanical way, neglecting the intrinsic connection of knowledge points [6]. And the knowledge mapping which is aimed at discovering the historical evolution process of scientific theories and methods can be found by mining and analyzing the knowledge mapping relationships to discover: discipline structure characteristics, research hotspots, development sources, professional relevance and breakthrough achievements, and future development directions [7]. At present, informatization of teaching knowledge can help more people learn more knowledge, and students can not only learn book knowledge in the classroom but also learn some other knowledge through the Internet after class time, which can well help students check the gaps [8].

Nowadays, an emerging cross-learning knowledge mapping, which is based on knowledge units and enables effective access to knowledge and quick grasp of the frontier areas of disciplinary knowledge, is emerging [9]. Knowledge mapping is a kind of graph showing the process of knowledge development and its interstructural relationships based on content analysis, citation analysis, and visualization [10]. It can describe human-owned knowledge resources and their carriers over time using visualization techniques, map, mine, analyze, and display scientific knowledge and their interconnections, and create a knowledge-sharing environment within organizations to promote collaboration and deeper scientific and technological research [11]. Text CNN is a machine learning model under deep supervised learning, and compared with traditional machine learning for manual feature extraction, Text CNN can automatically extract local features with shared weights, which is better than traditional machine learning algorithms for text classification. From the initial study of discrete models of TSP to the present, several generations of combinatorial mathematicians and operational researchers have studied this problem extensively and intensively and have achieved great achievements. Using the same global search strategy-based algorithm to train the neural network parameters can avoid the above-mentioned drawbacks of the genetic algorithm, because its velocity-a-displacement model is simple to operate and only decides the search direction according to its own velocity.

This paper combines quantitative analysis with qualitative analysis and introduces the theory and method of scientific knowledge mapping based on scientometrics. Scientometrics studies Maritime English and draws a series of knowledge maps of Maritime English, which is a new attempt to study the knowledge system in this field. The innovative contribution lies in the attempt to use the core and edge structure analysis function options in word2vec software to visualize the social network map of Maritime English researchers in the process of using social network analysis method to build knowledge map. This is a pioneering exploration on the technical details of knowledge map construction. At the same time, in order to maintain the diversity of particle swarm optimization, Text CNN algorithm is combined with the construction of Maritime English subject knowledge map, and the network parameters and structure are optimized at the same time. It is of practical significance and academic value to make use of knowledge map to study Chinese Maritime English, which is helpful to the development of Chinese Maritime English and provides guidance.

2. Thoughts on Constructing Knowledge Map of Maritime English Based on Text CNN Algorithm

2.1. Construction Framework of Subject Knowledge Map Based on Text CNN Algorithm. The disciplinary knowledge mapping constructed in this paper is guided by the disciplinary ontology and takes the information of disciplinary entities and the relationship between entities as the core to construct the knowledge mapping within the scope of the disciplinary domain. At present, knowledge mapping is still limited to specific disciplines or fields, and studies are conducted for the knowledge contents within the disciplines or fields to meet the needs of people in specific professional fields and to conduct accurate and refined search for professional academic issues. It can display the tedious knowledge in the field through information processing, data mining, graphic drawing, and knowledge measurement and provide valuable and practical reference for discipline research to reveal the knowledge areas with dynamic development laws. The framework for the construction of disciplinary knowledge mapping is shown in Figure 1.

First, the extraction of subject entities and the relationships between entities are divided into data preprocessing, acquisition of labeled data, word vector representation, feature extraction, and relationship classification for implementation. The forward neural network function class N is dense in the space of vector-valued continuous functions, in a consistent paradigm sense, where the forward network function class is defined as a set of the following types.

$$\mathrm{N} = \left\{ \mathrm{Y} \in \mathrm{R}^{\mathrm{m}}, \mathrm{Y} = \mathrm{W}^{(\mathrm{K}+1)} \mathrm{N}_{\mathrm{k}} \{ \mathrm{N}_{\mathrm{k-1}} [\cdots N_1(X)] \} \right\}, K = 1, 2, 3 \cdots \tag{1}$$

CiteSpace is used as a construction tool, combined with scientometrics theories and methods such as word frequency analysis method and cooccurrence analysis method, to study the knowledge map of Chinese maritime English subject research subjects. According to the characteristics of the large sample data set, the Bayesian method is used to extract the prior information, which is coupled into the training neural network. Its input is W_{x}, output is x; then the neural network is:

$$\frac{dx}{dt} = -WF'(x)^T F(x). \tag{2}$$

Suitable data objects in the field of Nautical English are selected and used as data sources for knowledge graph construction; then normalized preprocessing is carried out, including data text format conversion, Chinese word separation, deactivation word list filtering, indexing, data feature item extraction, and data annotation. We study and use theories, construction methods, and mainstream tools related to disciplinary knowledge mapping: CiteSpace, VOSviewer, etc., and functional tools Bibexcel (mainly for foreign language data text merging, data format conversion, field extraction and weighting, cooccurrence frequency calculation, and cooccurrence matrix generation), SATI (mainly for Chinese title field extraction, frequency statistics, and matrix generation), and Bicomb (mainly for bibliographic field extraction, frequency statistics, matrix generation). The first N words with greater weight are selected for output to get the concept set of the disciplinary knowledge map and to

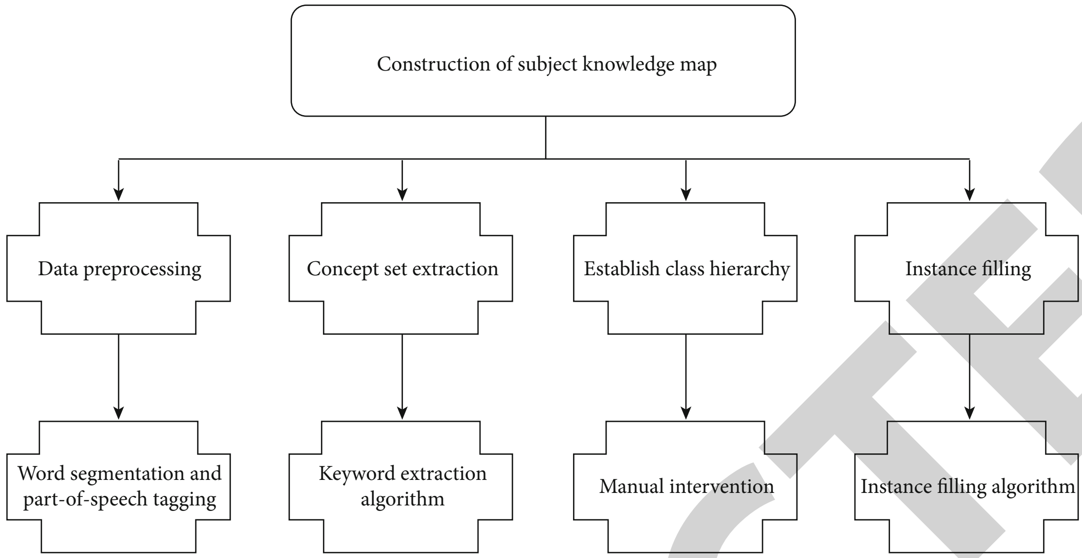

Figure 1: Framework of discipline knowledge map construction.

do the corresponding groundwork for constructing the disciplinary ontology later. The keyword extraction process is shown in Figure 2.

Second, data preprocessing is performed for the extraction of entities and their relationships in the subject area, that is, determining the scope of disciplinary knowledge graph construction-computer network disciplines, as well as performing manual culling, clause splitting, and other operations for computer network disciplinary texts. Using the interlayer local spatial correlation to connect the neuron node of each adjacent layer only to the upper layer neuron node that is close to it, i.e., local connection, a multilayer forward network is built with the function to calculate the output M of the hidden layer:

$$M_j = f\left(\sum_{i=1}^{n} W_{ij}X_i - b_j\right), j = 1, 2, \cdots, m. \quad (3)$$

The data layer is an ontology-related data enrichment process that emphasizes the relationships between instances of English disciplines and their attribute-value pairs. Latent semantic algorithm analysis is performed on the data after preprocessing to obtain the results of the processed data. The two functional features extracted are upper and lower bounds and first-order derivatives, which are transformed into equivalent mathematical expression embeddings for training the neural network. Then the number of nodes is increased sequentially until the recorded error value is minimized. The formula for determining the number of nodes in the hidden layer using the trial-and-error method is as follows:

$$m = \sqrt{k + l} + a, \quad (4)$$

where m is the number of hidden layer nodes, k is the input node, and l is the output node.

In conjunction with the theoretical knowledge of discipline construction, the initial steps of constructing a disciplinary knowledge mapping tool reveal a complete picture of the practical and cognitive aspects of the discipline. The core is to identify and correlate the mainstream tools with the conceived disciplinary knowledge map construction scheme and iterate on them to improve the scheme step by step. The theoretical support and model of the disciplinary knowledge mapping scheme are also proposed.

Finally, the subject entities are obtained using hanlp, the labeled data with the relationship information between the entities are obtained using remote supervision, and the obtained corpus with the labeled information is used as the input of the neural network model. The relationship classification model is trained so that the model can automatically discriminate the relationship of entities in a sentence. The results of the data analysis obtained are used for visual technical mapping to obtain the disciplinary knowledge map. The feasibility and practicality of the constructed scheme are verified and optimized through the discipline of Nautical English, especially the construction method fusion. Relying on Text CNN to automatically extract features from textual information represented at the simple character level, the feature vectors are fed into the LSTM model via the highway network framework based on sequences thus achieving better classification prediction. Among them, the convolutional layer can well characterize the local features of the input data, while the pooling layer can further extract the most important part of the local features on this basis.

2.2. Entity Relationship Extraction Method. The most basic as well as the key to constructing a knowledge map within the subject area is to obtain the information of entities within the subject area and to realize the extraction of relationships between entities. Maritime English reading is a professional course with strong theoretical and practical aspects, which requires students to have a good professional basic knowledge before learning. The strong foreign-related

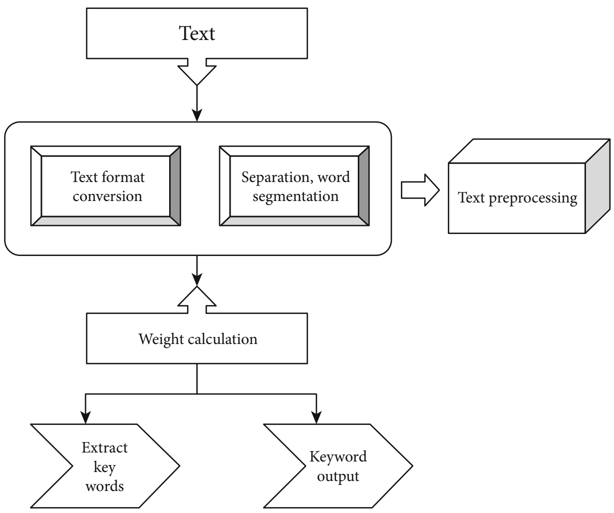

Figure 2: Keyword extraction process.

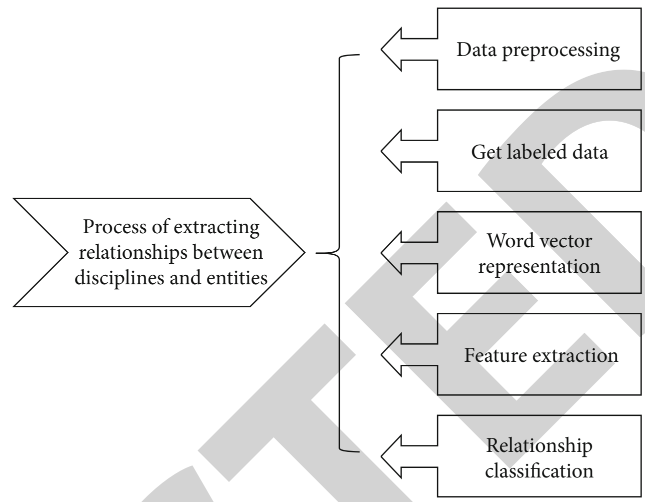

Figure 3: Process of extracting the relationship between subject entities and entities.

characteristics of the nautical profession and the strong professionalism of the nautical professional listening and speaking courses determine that the nautical professional listening and speaking teaching is different from the teaching methods of other courses. The convolutional neural network structure has been proved to be very effective in such tasks as image recognition and speech recognition. The process of subject entity and interentity relationship extraction is shown in Figure 3.

First of all, in the field of education, there are different subject entities for each subject, taking computer networks as an example, such as protocol names and device names, called named entities. However, in the current teaching of maritime English, due to the assessment method, the same paper-based examination as other courses is used. Therefore, many courses are set up to focus on how to improve students' reading ability, ignoring the training of speaking, listening, and writing ability, which leads to students' inability to understand and speak after they join the workforce. Therefore, the Chinese data literature catalogues are screened and saved in endnote format that SATI software can recognize, output as data files with txt format, and imported into SATI, and through data conversion processing, the similarity matrix and related data tables that can be used for analysis are generated for subsequent analysis. In the later stage of the search, the discovery probability is reduced in order to increase the convergence speed of the algorithm. Therefore, the discovery probability is improved as follows:

$$\mathrm{pa}_t = \exp\left(t/t_{\max}\right) \times \cos\left(t/t_{\max}\right) \times \mathrm{pa}_{\mathrm{begin}}, \tag{5}$$

where pa_t is the discovery probability of t iteration and $\exp\left(t/t_{\max}\right) \times \cos\left(t/t_{\max}\right)$ is the function dynamic decreasing factor.

The subsequent max-pooling layer allows for the secondary extraction of the most expressive features in each feature map. In order to implement a pure character-level embedded convolutional neural network, the first thing to do is to construct the alphabet. The alphabet used in this paper is as follows, with 79 characters, for which one-hot encoding is used, plus an all-zero vector (for characters not in the character list, including unknown and null characters). The N-gram feature is a sequence model of every adjacent n bytes in a given malicious sample, which can be understood as a first-order Markov chain, that is, the order of occurrence of byte sequence is a random process of discrete events. By calculating the correlation between different granularity N-gram features, the different granularity features are correlated with each other, the corresponding weights are assigned to the different granularity features, and their representations are obtained by weighted summation:

$$\alpha_{\mathrm{i}}^t = \frac{\exp\left(U_i^{t^T} u_w\right)}{\sum_t \exp\left(U_i^{t^T} u_w\right)}, \tag{6}$$

where c_i^t is the i word representation of t feature.

Second, the extracted keywords are used to align the existing remote knowledge base and extract the corresponding interentity relationship information to obtain the subject-wide entities, relationships, and entity triads. English teachers in maritime colleges and universities often do not have enough professional knowledge; thus, they are not able to teach maritime English. Therefore, the similarity matrix of high-frequency words is obtained by using the formula of low-frequency word boundaries of high-frequency words, that is, the matrix with a diagonal of 1. The data in the table indicates the frequency of any two keywords appearing together in the same literature, and the larger the value indicates the greater the degree of correlation or similarity between these two keywords. If the total input of the jth neuron N_{j} in the network is defined as u_j and the output state is v_j, then the state transfer equation of the network can be written as:

$$v_j = g(u_j) = g\left(\sum_t w_{ij} x_j - q_j\right), \quad (7)$$

where g is an S-type function, commonly used as:

$$g_1(x) = \frac{1}{\left(1 + e^{-\lambda x}\right)}, \quad (8)$$

$$g_2(x) = th(\lambda x). \quad (9)$$

Next, Chinese pinyin sequences are introduced to semantically expand the original text, the Chinese text is converted by hanyu pinyin, and the pinyin-represented text is then embedded by character. After multilayer convolution and pooling operations, the obtained feature maps are expanded sequentially by rows, connected into vectors, and input into a fully connected network, which usually uses Softmax logistic regression as the feature classifier. As the large-scale search gradually stabilizes and the update speed slows down, the coefficients of the speed $\mathrm{v}_\mathrm{i}^\mathrm{t}$ at the t moment are adjusted by using the property that the inverse cosine function is monotonically decreasing on $[0, 1]$ as follows:

$$\begin{aligned} \mathrm{w}(t) = sqrt[4/7 \times \arccos\,(t/t - \max)] \\ \times [(t - \max - t)/(t - \max - 1)], \end{aligned} \quad (10)$$

where $w(t)$ is called the inverse cosine function speed adjustment factor. The speed update formula is transformed into:

$$v_i^{t+1} = w(t) \times v_i^t + \left(x_i^t - x^*\right) F_i. \quad (11)$$

Finally, the entity pairs extracted from the textual information are aligned to the remote knowledge base, and the remote knowledge base is used to automate the annotation of the relationships between entities, or "transmission units," to achieve the extraction of triples in the subject area, in order to automate the extraction of subject knowledge using neural networks. Therefore, it is necessary to convert the input statements into a form that can be recognized by the machine, i.e., a vectorized representation. Due to the inevitability of differences in experiential background (e.g., in terms of background knowledge of sailing acquired by students from coastal and inland regions), learners' perceptions and understanding of problems are often very different. So, in order to ensure that the whole classification model does not have training difficulties due to excessive depth, the feature extraction stage does not use the stacked convolutional layers to extract high-dimensional sufficiently abstract features, but a novel framework of a single-layer Text CNN plus highway network is used instead. It refers to the fact that the nodes of the convolutional layer are only connected to some of the nodes of the previous layer and used to learn local features. This connection significantly reduces the number of parameters, speeds up the learning efficiency, and reduces the possibility of overfitting to some extent.

3. Application and Analysis of Text CNN Algorithm in the Construction of Knowledge Map of Maritime English

3.1. Training Analysis of Text CNN Algorithm. Up to now, the method used for training convolutional neural networks is still the traditional gradient descent method. Among them, if the batch gradient descent method is used, although the best convergence effect can be obtained, the convergence speed of the training process is severely limited because all training samples are required to participate in the operation in each iteration process. The advantage of Bi LSTM is that it can learn the dependence between observation sequences (input words) through two-way settings. In the training process, LSTM can automatically extract the features of observation sequences according to targets (such as recognition entities). Therefore, the Text CNN algorithm uses the Bi LSTM to process the input vector, and then at each time step, the output of the Bi LSTM is stitched with the corresponding word vector as the "semantic vector" of the current time step, which can well represent the contextual features of the text. The accuracy and loss curves of the dataset during the training process are shown in Figures 4 and 5.

First, each iteration process requires only a small number of samples to participate, and shipping samples can speed up the convergence while ensuring that the optimal solution is found as much as possible. In the training phase of the algorithm model, the input is labeled data, and the label of the document is the output. Planes of the same depth are called depth slices, and the same slice shares the same set of weights and biases. The repetition unit is able to identify both the feature, without considering its position in the viewable domain, helping the neural network to remain spatially invariant to the input. Individual genes are evaluated according to some criterion, resulting in a ranking table of importance for each gene, and then the gene with relatively high evaluation score is selected as the feature gene based on the ranking result. The attributes are conditionally independent of each other for a given target value. In order to avoid the information carried by other attributes being used by attribute values that never appear in the training set, the probability values are usually modified when estimating. Laplace correction is often used. The assumption of attribute conditional independence is relaxed to some extent. Directed acyclic graph is used to describe the dependency relationship between attributes, and conditional probability table is used to describe the joint probability distribution of attributes. The advantage is that few parameters are required to be estimated and less sensitive to missing data. In order to improve the accuracy of the network parameters, the number of iterations has to be increased, and then the computational effort is also greatly increased. The size of the response of the Text CNN network to the input depends on the distance between the input vector and the center of the network, and the smaller the distance between the input vector and the center, the larger the response of the neuron. So the center correction process of Text CNN is essentially the process of clustering the input samples based on the distance between them, the input vectors with small distances from

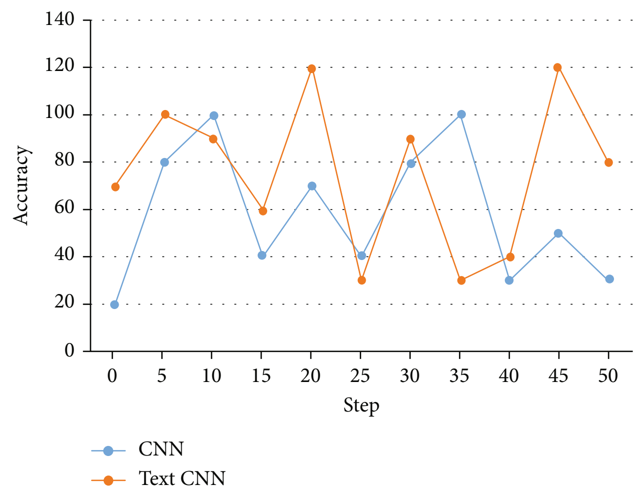

Figure 4: Curve of accuracy rate changing with time during training.

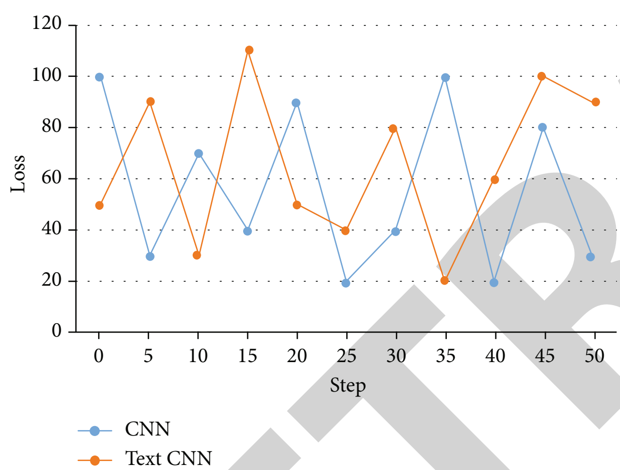

Figure 5: Curve of loss with time during training.

each other are grouped into one class, and the center of clustering is the network center. Three datasets of different sizes are selected for comparison experiments, and the details are shown in Table 1.

Second, to prevent overfitting in the training of the model, L2 regularization is used to constrain the parameters of the convolutional neural network. Some kind of metric is usually used to evaluate the importance or relevance of each gene to the classification, then the genes are ranked according to their importance, and finally, a certain number of top-ranked genes are selected as the feature genes; among such methods, the signal-to-noise ratio evaluation metric is the most commonly used. Corresponding initial search instructions should be made to form a unique search format by example queries, so that users can find the queried data more easily and quickly. Logistic regression is a generalized linear regression analysis model, which is used to deal with regression problems where the dependent variable is a categorical variable and is commonly used for dichotomous or binomial distribution problems, where the graph of the relationship between the dichotomous probability and the independent variable is mostly an *S*-shaped curve and can also deal with multicategorical problems. Based on the tensor flow deep learning platform, using one GPU and trained with a convolutional kernel width of 3, the results of Google Billion Word benchmark test of Text CNN trained on a single GPU are shown in Figure 6.

Table 1: Details of data sets of different sizes.

Data	Number of target classes	Average length of text	The maximum length of the text
MR	3	26	28
SST	5	45	65
Irony	2	37	84

Finally, a dropout strategy is introduced for training the final fully connected layer parameters, i.e., a portion of the trained parameters is randomly selected for discarding at each update. In the hierarchical clustering algorithm, after the region division is completed in the lower layer, the path-finding optimization of each region in the higher layer actually uses only the center of gravity coordinates of the neighboring regions, and this information is obtained in the process of region division in the lower layer. In addition, the similarity matching process should be done, the queried features should be compared with the feature values in the database through certain algorithms, and the data can be returned to the user only after satisfying certain similarities to ensure that the user's retrieval needs can be met. Representing instances as points in space, the mapping makes instances of separate categories separated by distinct intervals as wide as possible, maps the new instances to the same space, and predicts the category to which they belong based on which side of the interval they fall. After that, make the number of hidden layer nodes plus one and still go through a certain number of gradient descent iterations as in the beginning, if after the iterations the objective function decreases by a value greater than the threshold value, this means that an additional hidden node has been added, which is more useful for the network, and the number of hidden layer nodes at this time does not make the objective function of the system extremely small, so make the number of hidden layer nodes plus one and continue the iterations as above.

3.2. Feature Extraction Analysis of Text CNN Algorithm in Subject Knowledge Map. The ultimate goal of teaching conversational maritime English is to equip students with the ability to communicate effectively in English in real work situations. Knowledge graphs are usually expressed using the semantic technology standard language RDF or ontology language, and the construction of domain knowledge graphs is often based on the concept of domain knowledge ontology. Therefore Text CNN algorithm is mainly to extract the content needed by users by processing the original media data, and the process is more focused on the accuracy of information retrieval. Then it needs to be

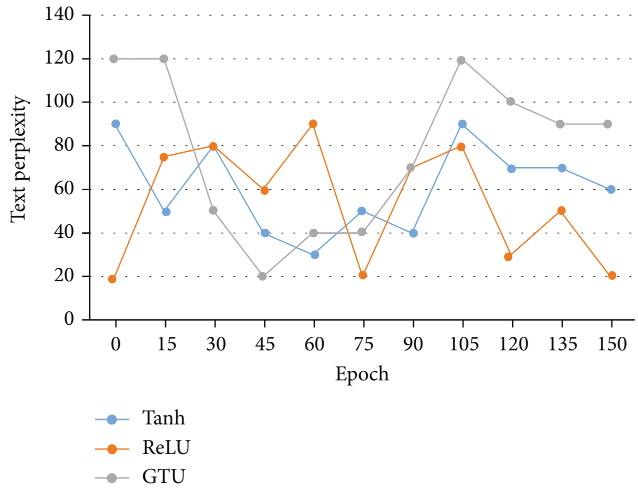

Figure 6: PPL change curve during training.

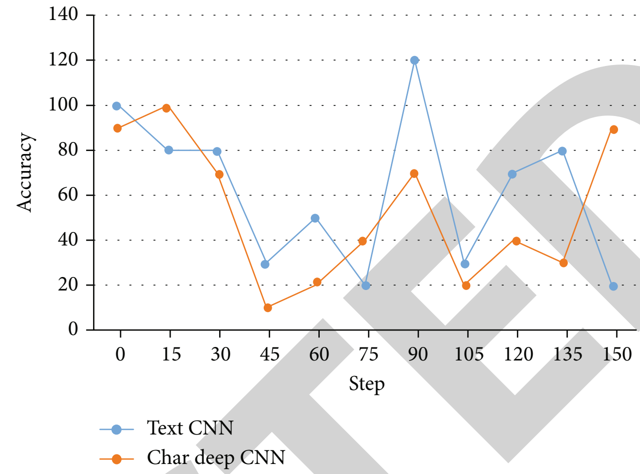

Figure 7: Curve of accuracy during training.

trained with data with category labels, and each sample data contains not only a number of features, i.e., the sample expression values in all genes, but also the category to which it belongs. The shipping makes the data itself already have better distribution characteristics. The classification performance on the English dataset will be compared with several common deep learning models in terms of accuracy. A comparison of the accuracy change curve between the Text CNN algorithm and Char Deep CNN during the training process is shown in Figure 7.

First, in a certain document, a word or phrase appears repeatedly, i.e., the word frequency TF is high, while distinguishing from other documents, the word rarely appears in other documents, i.e., the word frequency TF is low, which means that the word or phrase is more representative of the features of the document and can be used as feature extraction for classification. Training word2vec in advance, using it as input features for CNN, and updating it continuously during iterative training, shipping is equivalent to introducing certain prior knowledge that can guide the model to converge to the optimal solution in a better direction during the training process. It usually employs a classifier to directly evaluate the classification performance of a subset of feature genes and then adopts some strategy to adjust the subset according to the evaluation results to achieve the purpose of continuously exploring the optimal subset. Coword clustering analysis is to gather the ones with high similarity of subject words together to form a category with high intragroup similarity, low intergroup similarity, and relatively independent concepts. The existing methods for evaluating text classification are mainly judged by the model's prediction of correct text labels. Table 2 shows the mixture matrix built according to the Text CNN algorithm, which is used to introduce the calculation of evaluation metrics.

Second, text representation using vector space models requires a lot of work to annotate the text if lexical features need to be preserved. In contrast, word2vec simplifies the manual work by converting words in the text into vector representations based on contextual features. Feature extraction and model training are not completely separate, and the input distributed features are updated as parameters during the iterative training process. The performance of the classification is tested on the whole dataset, and then the change in performance after subtracting each gene is calculated. The gene with the smallest absolute value of association weight in the classification function is selected and removed from the training set, the process is repeated until the training set data is empty, and the subset of feature genes removed in the last step is the optimal classification subset. A trifold cross-validation method is used to select training and test samples, which allows the classification model to be fully learned. The relationship between Text CNN and SVM in terms of feature dimension and classification error rate is shown in Figure 8.

Table 2: Correlation of evaluation indicators.

	Predict	Not retrieved
Real	16	5
Irrelevance	31	7
Grand total	47	12

In the actual entity relationship extraction process, its individual sentence lengths are often inconsistent. It also means that this paper needs to extract the inconsistent local features from the individual sentences with inconsistent lengths for predicting the relationship types of the target entities. Using this method to train the network not only shows very intuitively and clearly the magnitude of the influence of each hidden layer central node on the deviation decline rate of the network. Moreover, the number of iterations is also very small, and generally only a few iterations are needed to finalize how many hidden layer nodes are needed.

Finally, for the analysis of large amounts of text data, we need a tool to represent the text as data understood by the computer; in other words, word2vec is a tool to convert text into a numerical representation. The skip-gram model for feature extraction and the convolutional neural network model for classification should be considered whole, and

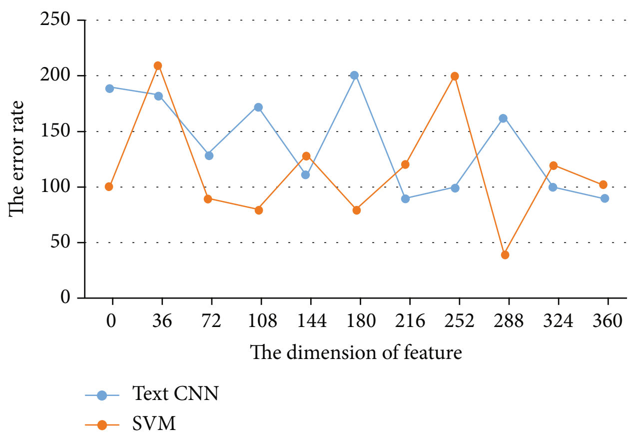

Figure 8: Relationship between feature dimension and classification error rate.

together, they perform the whole process of the short-text sentiment classification task. The recursive hierarchical feature gene selection method is repeatedly run according to different training sample distribution structures, guided by the classification accuracy, and then, the final selected subset of feature genes is obtained by synthetically integrating numerous feature selectors. The maximum pooling operation is used to perform secondary extraction of features, that is, for each filter set in the convolutional layer, the value of this paper selects the maximum of them as the final retained value, and the rest is discarded. The network accuracy is further improved with each additional network center, and the next center selected each time is the one that contributes most to the error reduction in the remaining part.

4. Conclusions

This paper presents a method of constructing Maritime English subject knowledge map based on Text CNN algorithm. The application of Text CNN algorithm in the construction of subject knowledge map is analyzed. It makes knowledge more relevant and hierarchical. At the same time, the knowledge retrieval technology of knowledge mapping realizes the accurate answer to relevant English questions and can effectively help students check and fill in the gaps of English knowledge points. In the process of using social network analysis method to build knowledge map, try to use the core and edge structure analysis function options in the word2vec software to visualize the social network map of Maritime English researchers. In order to maintain the diversity of particle swarm optimization, Text CNN algorithm is combined with the construction of Maritime English subject knowledge map, and the network parameters and structure are optimized. The knowledge map construction method based on text CNN algorithm proposed in this paper can reveal the English subject knowledge system and standardize the knowledge structure. It is of practical significance and academic value to make use of knowledge map to study Chinese Maritime English, which is helpful to the development of Chinese Maritime English and provides guidance.

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 known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was financially supported by the Education and Teaching Reform Research Foundation of Higher Education Institution of Hainan Province of China: Evaluation of Nautical English Ability Based on Network Learning Platform and Teaching Improvement Research (Grant No. Hnjg2021-117) and the Hainan Province Philosophy and Social Science Planning Project: A Study on the International Ship Registration System under the Background of Free Trade Port (Grant No. HNSK(ZX)21-90).

References

[1] Q. Kuo and X. Shijie, “The application of subject knowledge map in constructing middle school English curriculum map,” *Modern Communication*, vol. 18, p. 2, 2017.

[2] C. Hui, D. Hui, and W. Daoting, “Research on automatic construction of knowledge graph for instructional design discipline based on natural language processing,” *China Education Informatization*, vol. 7, p. 5, 2020.

[3] X. Zhu and Z. Xie, “Automatic knowledge graph construction algorithm based on massive text data,” *Journal of Jilin University: Engineering Edition*, vol. 51, no. 4, p. 6, 2021.

[4] X. Wu, “The current situation and countermeasures of China’s maritime English teaching research,” *Maritime Education Research*, vol. 35, no. 1, p. 7, 2018.

[5] J. Wang, “On the design and application of digital teaching resources in navigational English listening and conversation,” *Journal of Jiamusi Vocational College*, vol. 194, no. 1, pp. 212-213, 2019.

[6] Z. Xiaolan, “Discussion on design elements of navigation English digital teaching resources,” *Journal of Shandong Agricultural Management Cadre College*, vol. 35, no. 4, pp. 120-121, 2018.

[7] A. Nugroho and I. Mutiaraningrum, “EFL teachers’ beliefs and practices about digital learning of English,” *EduLite: Journal of English Education, Literature and Culture*, vol. 5, no. 2, pp. 304–321, 2020.

[8] D. Hongxin and L. Jiaxiang, “Analysis of the application of structural metaphor cognition in vocabulary teaching and learning of Nautical English,” *Caizhi*, vol. 1, p. 1, 2020.

[9] Z. Xiaolan, “Constructing a digital learning platform for professional English——taking marine English as an example,” *Journal of Shandong Agricultural Engineering Institute*, vol. 34, no. 10, p. 3, 2017.

[10] Y. Xu, “Research on the teaching reform of sailing English under the background of "internet +",” *Journal of Zhejiang International Maritime Vocational and Technical College*, vol. 13, no. 1, p. 3, 2017.

[11] Y. Chen and Y. Lin, “Impact of smart cities on resource conservation based on big data analysis,” in *Data Processing Techniques and Applications for Cyber-Physical Systems (DPTA 2019). Advances in Intelligent Systems and Computing*, C. Huang, Y. W. Chan, and N. Yen, Eds., vol. 1088, Springer, Singapore, 2020.