Hot Keyword Extraction of Sci-tech Periodicals Based on the Improved BERT Model

Bing Liu[1*], Zhijun Lv[1], Nan Zhu[1], Dongyu Chang[1], and Mengxin Lu[1]
[1] School of Economics and Management, Dalian University of Technology
Dalian, CA 116024 China
[e-mail: liubingbs@163.com]
[*]Corresponding author: Bing Liu

Received February 19, 2022; revised April 28, 2022; accepted May 25, 2022; published June 30, 2022

Abstract

With the development of the economy and the improvement of living standards, the hot issues in the subject area have become the main research direction, and the mining of the hot issues in the subject currently has problems such as a large amount of data and a complex algorithm structure. Therefore, in response to this problem, this study proposes a method for extracting hot keywords in scientific journals based on the improved BERT model.It can also provide reference for researchers,and the research method improves the overall similarity measure of the ensemble,introducing compound keyword word density, combining word segmentation, word sense set distance, and density clustering to construct an improved BERT framework, establish a composite keyword heat analysis model based on I-BERT framework.Taking the 14420 articles published in 21 kinds of social science management periodicals collected by CNKI(China National Knowledge Infrastructure) in 2017-2019 as the experimental data, the superiority of the proposed method is verified by the data of word spacing, class spacing, extraction accuracy and recall of hot keywords. In the experimental process of this research, it can be found that the method proposed in this paper has a higher accuracy than other methods in extracting hot keywords, which can ensure the timeliness and accuracy of scientific journals in capturing hot topics in the discipline, and finally pass Use information technology to master popular key words.

Keywords: Bidirectional encoder, hot keyword, representations from transformers (bert), sci-tech periodicals, similarity measurement.

1. Introduction

As an important carrier for recording and disseminating new scientific discoveries, sci-tech periodicals are an important platform for scientific and technological exchanges[1]. It plays an irreplaceable role in promoting the development of science and technology. The role orientation of sci-tech periodicals is changing to enlightening thinking, improving innovation ability and leading the development of disciplines[2]. Follow up academic trends, plan forward-looking and leading topics, timely open up academic columns of scientific and technological periodicals, publish annual topic selection guidelines, hold special academic conferences, organize targeted academic activities, publish special academic periodicals related to sci-tech periodicals, exchange and report on the cutting-edge direction of discipline development, are important ways for sci-tech journal to lead the development of disciplines[3-4]. How to grasp the hot topics of discipline research in time, dynamically adjust the topic selection direction and development strategy of periodicals, and continuously improve the academic influence of periodicals has become a hot topic in periodicals in recent years[5].

Under the traditional publishing model, the planning and organization of publication activities such as manuscript assembly, review and promotion of periodicals mainly relies on the experience and market intuition of journal managers, resulting in a serious lag in timeliness and insufficient accuracy in hot spot extraction[6]. At present, in journal planning, big data technology can be used to realize data mining, and deep learning tools can be used to discover research hotspots and realize intelligent management of journal dynamics. Therefore, this study can use big data technology and deep learning to mine the hot keywords of journals and discover potential rules.

2. The Research Status

At present, the widely used method is to conduct relevant research on subject hot issues mining based on keywords. For the analysis of keywords, the word frequency analysis method is usually used to determine the research hot issues and their changes by detecting the frequency of keywords in a research field.

Chu [7] collected the keywords of research papers relevant to knowledge management on the CSSCI database from 2000 to 2009, and recognized research hotspots in the field by analyzing the distribution of high-frequency keywords. Zhao[8] used the average year of occurrence obtained by high-frequency keyword clustering to observe research hotspots. On the basis of word frequency analysis, Atlam [9] estimated the change of keyword research popularity in a certain time domain with machine learning method. Li[10] proposed a concept of evolving co-word network, and analyzed the research hotspots in the field of life energy through cluster analysis of co-words. Luo[11] proposed the concept of dynamic co-word network, and applied it to the study of hot issues in the field of soil heavy metal pollution bio-remediation. Wang[12] used Bibexcel, Ucinet, and Netdraw to obtain high-frequency keywords and identified research hotspots through cluster analysis by SPSS. The existing research method is mainly to study the change of the frequency of keywords in periodicals. But most of the keywords in periodicals are compound words with rich meanings composed

of multiple simple words. A single compound word vector is difficult to effectively express the differences in different information elements.

The above-mentioned methods mainly discover research hotspots based on the change of keyword frequency, but keywords are mostly compound words with rich meanings which are composed of multiple simple words, and a single compound word vector cannot effectively express the differences in different information elements. Word frequency analysis directly using journal keywords will cause a single compound word vector to be sparse in clustering, and it is difficult to obtain accurate distribution results. At the same time, in the research of popularity analysis, there are fewer subdivisions of domain description words, which makes the frequency of domain words higher than that of technical words. In summary, the traditional keyword frequency analysis can reveal the research hotspots and development trends of the research area to a certain extent, but the lack of word frequency statistics of the language model makes it difficult to further accurately measure the similarity of word meanings. Especially for the research direction of early development, when the expression of keywords has not been unified and the word meaning model has not been formed, it is difficult to accurately count the real frequency of words, leading to the delay of hotspot discovery. Journal topic planning needs to further discover and categorize keywords for thousands of papers. The core operation is to find multiple hotspot centers from the keyword distribution, and further realize more knowledge mining.

With the development of natural language processing technology, new text mining technologies are emerging. The TFIDF method extracts keywords by counting the word frequency of the large corpus and the current text word frequency[13]. The Textrank method judges whether a word is a keyword from the co-occurrence weight, but it cannot achieve semantic measurement[14]. With the emergence of word vector technology that converts words into numerical vectors, word meaning measurement becomes possible. The main derived word vector generation models include Word2vec[15], G1oVe[16], ELMo[17], and BERT[18]. The most commonly used are the Word2vec model and the BERT model. The Word2vec model extracts the weight matrix in the middle of the network structure as a vector dictionary, which makes some words unable to obtain vectors and can not deal with more diverse compound words. The BERT model is implemented by the Transformer encoder, which can analyze contextual semantics, accurately quantify polysemous words, and acquire vectors of compound words of any length[19]. It is currently one of the most effective language models in natural language processing tasks. Rodriguez et al. [20] conducted in-depth research on machine learning and data mining methods, theoretically analyzed the practicability of machine learning and data mining methods in text keyword extraction, and preliminarily used the mining algorithm in this paper to conduct the experimental analysis. Mohammadi Ehsan et al.[21]thought about text mining, verified the effectiveness of text mining algorithm for journal data processing, and verified that text recognition algorithm has better effect in similar data processing methods through experimental analysis.Bayrak Tuncay[22] has studied the text mining algorithm, and believes that text mining and big data should be integrated with each other, and the effect of text mining should be improved by using big data technology.

According to the current research situation, the traditional keyword frequency analysis can reveal the research hotspots and development trends of the subject to a certain extent. But it is

difficult to further accurately measure the similarity of word meaning due to the lack of word frequency statistics of language models. Especially for the early development of the research direction, the expression of keywords has not yet formed a unified word meaning model. It is difficult to accurately count the frequency of hot words, resulting in delays in the discovery of hot spots. Therefore, this paper takes the BERT model as the basis of word vector acquisition, divides the document compound keywords into word vector set expression, and retains the compound word vector in the set; at the same time, it innovatively defines the set similarity measurement method so that each element information can participate in the measurement independently, and retains the compound characteristics of the compound word vector. Finally, the improved BERT (i-BERT) composite keyword popularity analysis model is constructed. The popularity keywords are obtained according to the order and frequency center of the keywords. Taking 14420 management journal literature as the sample data, the advantages of the proposed method in hot keyword extraction are verified by comparing with TF-IDF, Word2vec and BERT methods.

3. Model Construction

In order to make the description of keywords more accurate, the keywords involved are often compound phrases composed of several simple words, such as coordinate phrases, subordinate phrases, subject-predicate phrases, verb-object phrases and so on. For example, the compound word AB is composed of two simple words A/B, or the compound word CDE is composed of three simple words C/D/E. In order to obtain the complete vector of this kind of phrase, the sentence vector acquisition method of the BERT model that can obtain the word vector of any length is adopted. However, compound words often have the modification of several simple words. If the vector of compound words is obtained directly, the multi-element information will be mixed into a high-dimensional word vector. It is difficult to distinguish information elements of different dimensions when calculating word vector similarity. Considering that simple words have information independence when compound words are combined, this paper further splits compound keywords, and uses a multi-element semantic set containing simple words and compound words as the word meaning expression of compound keywords. The vector of elements in the set adopts the BERT word vector model, and strengthens the weight of consistent information in the similarity measurement, making the aggregation of compound words no longer uniform, while synonyms attract each other. Finally, self-generating tags of keywords in the class are realized, so that the generated central word has a certain generality for the class and retains its original information. The process of building the model is shown in **Fig. 1**.

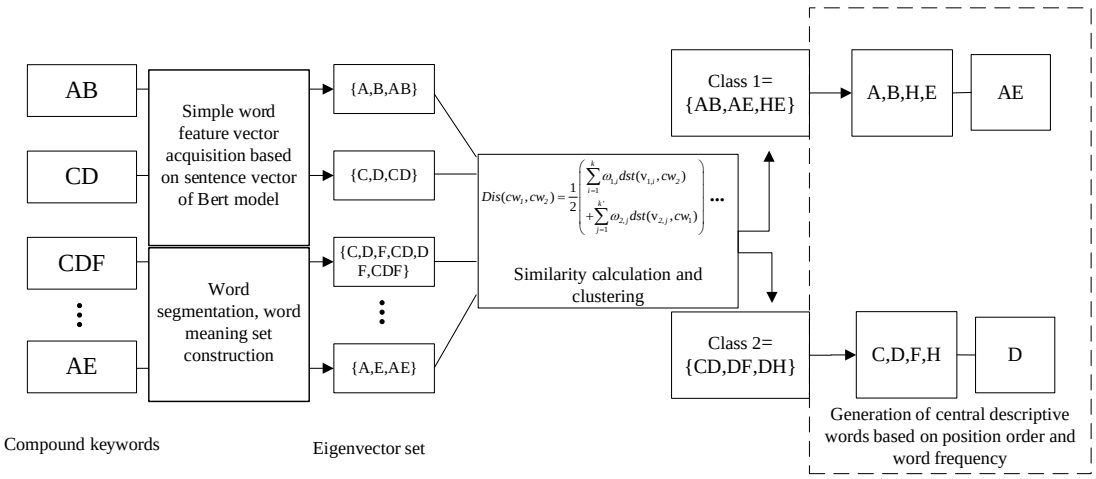

Fig. 1. Flow chart of model construction

Step1: Perform word segmentation on all j compound keyword sets $mk_i = \{cw_1, cw_2, \cdots, cw_j\}$ of all periodicals sample sets in the i-th month of the sample set to obtain fuzzy word meaning sets, such as: keyword $cw_j = \text{'AB'}$ can be split into a fuzzy word meaning set $\{\text{A,B,AB}\}$.

Step2: The vector expression of each element in the set is obtained based on the BERT model, so that each compound word can be further represented as a split word meaning set $cw = \{v_1, v_2, ..., v_k\}$.

Step3: Determine the method of compound word set distance measurement, and combine CFSFDP word density clustering with density distance decision graph to obtain journal keyword class.

Step4: Split the keywords in the class to obtain the segmented word set, recognize the density center of the segmented word set in each set, and use the distribution divergence in the set as the confidence to decide whether to use the central word. All the retained central words are spliced in order to obtain the central words of the key word set.

At the same time, the maximum information coefficient can solve the relationship between attributes and targets in the word segmentation set. The influence degree of different attribute values on the target is estimated by calculating mutual information and entropy, and it is based on calculating functional dependence weights[23]. Therefore, each key word in the keyword set can be regarded as an attribute. The most frequently occurring keyword can be regarded as a target so as to calculate the probability that different key words occur frequently in the keyword, that is, the measure of the influence weight of the key word on the target.

Let X be the central word randomly obtained. The entropy of X can be expressed as formula (1) :

$$H(X) = -\sum_{x-X} p(x) \log P(X) \tag{1}$$

In Formula (1), $H(X)$ represents the amount of information. $P(X)$ represents the probability of influence. The bigger the $P(X)$ is, the bigger the $H(X)$ is.

In the keyword set, two key words are taken as a group. The criticality of different key words to each other is shown in Formula (2) :

$$H(X \mid Y) = \sum_{y-Y} P(y) \sum_{x-X} P(x \mid y) \log P(x \mid y) \tag{2}$$

In Formula (2), H(X | Y) represents the probability of Y's existence when X has a certain degree of influence.

Meanwhile, Y will also bring a certain loss value to X, as shown in Formula (3) :

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) \tag{3}$$

Formula (4) can be derived from Formula (2) and Formula (3) :

$$0 \le I(X;Y) \le \min\{H(X), H(Y)\} \tag{4}$$

Thus, the probability of frequent occurrence of different key words in keywords can be obtained, namely, the maximum information coefficient is:

$$SU_{MAX}(X;Y) = 2\left[\frac{I_{MAX}(X;Y)}{H(X_i) + H(Y)}\right] \tag{5}$$

3.1 Obtaining Word Vectors

After obtaining the central word of the keyword set, the vector of the word should be obtained because the vector can represent the semantics of the whole sentence. Assuming that all the ^{j} keyword sets of all sample periodicals in the i-th month are $mk_i = \{cw_1, cw_2, \cdots, cw_j\}$, the word vector after word segmentation can be expressed as a compound word vector set $cw_j = \{v_{j1}, v_{j2}, ..., v_{jk}\}$ composed of multiple simple word vectors. The Chinese word segmentation device in this paper uses the domestic open-source "Jieba" Chinese word segmentation tool to segment the compound word set $mk_i = \{cw_1, cw_2, \cdots, cw_j\}$. Each keyword in the set is generally divided into 1-3 basic simple Chinese words, and then combined into words of different lengths in order. For example, journal keyword “ABC” can be divided into {A,B,C,AB,BC,ABC}. Finally, the BERT model is used to convert each word in the set into a word vector format.

The important part of the BERT model is implemented based on the bidirectional Transformer encoder, and the model structure is shown in **Fig. 2**.

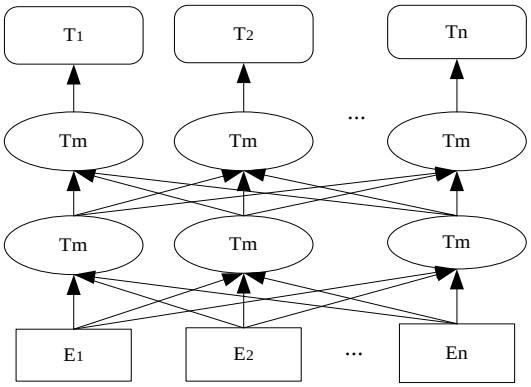

Fig. 2. Structure of BERT model

Where $E_1, E_2, \cdots, E_n$ represents the text input, and the vectorized representation of the text is obtained through the bidirectional transformer encoder, that is, the vectorized representation of the text is mainly realized through the Transformer encoder.

The output of the BERT model has two forms: one is the character level vector, that is, each character of the input short text has a vector representation; the other is the sentence level vector, that is, the leftmost [CLS] special symbol vector output by the BERT model, which thinks that this vector can represent the semantics of the entire sentence. This paper adopts the sentence vector output method to obtain the vector representation $cw_j = \{v_{j1}, v_{j2}, ..., v_{jk}\}$ for each set of compound keyword splits.

3.2 Keywords Extraction Based on the Improved Bert Model

Through experiments, it is found that in the word vector obtained based on the traditional BERT model, the similarity of two words can be divided into [same, very similar, general, completely dissimilar], and the corresponding Euclidean distance set is [0, 7.6, 10, 13.5], marked as $[0, \gamma_1, \gamma_2, \gamma_3]$. This reflects that in the measurement of semantic distance, the corresponding relationship between distance and semantics is not appropriate. Especially when two words are "the same" or "very similar", the semantic is very close, but the vector Euclidean distance $[0, \gamma_1]$ is significantly greater than the distance value $[\gamma_1, \gamma_3]$ between "very similar" and "completely dissimilar".

In order to solve this problem, this paper optimizes the distance of the traditional BERT model. The paper adopts the normalized nonlinear mapping of the distance weight of each pair of matching elements in the set so that the weight of similar words can be enlarged and the weight of unrelated words is at a lower value. The paper also establishes an improved BERT (i-BERT) model for the popularity analysis of the composite keywords so as to achieve the purpose of correcting the distance of the composite keyword set.

Optimization of compound keyword set distance: In the process of compound keyword set distance optimization, the first step is to calculate the similarity between two compound word sets $cw_1 = \{v_{11}, v_{12}, ..., v_{1k}\}$ and $cw_2 = \{v_{21}, v_{22}, ..., v_{2k'}\}$, which is interpreted as the sum of the bidirectional closest mapping distance between the elements of set cw_1 and set cw_2 under the

weight coefficient.The compound keyword set distance can be defined as:

$$Dis(cw_1, cw_2) = \frac{1}{2}\left(\begin{array}{l} \sum_{i=1}^{k} \omega_{1,i} dst(\mathrm{v}_{1,i}, cw_2) \\ +\sum_{j=1}^{k'} \omega_{2,j} dst(\mathrm{v}_{2,j}, cw_1) \end{array} \right) \tag{6}$$

Where

$$dis(v_{1,i}, cw_2) = \min\left[dis'(v_{1,i}, v_{2,1}), dis'(v_{1,i}, v_{2,2}), \cdots, dis'(v_{1,i}, v_{2,k'}) \right] \tag{7}$$

$$dis'(v_{1,i}, v_{2,1}) = (dis''(v_{1,i}, v_{2,1}))^2 \Big/ \gamma_1^{\,2} \tag{8}$$

In the above formula, $dis''(v_{1,i}, v_{2,1})$ is the Euclidean distance of the word vector in the BERT model. After the distance correction of compound keyword set, the original Euclidean distance $[0, 7.6, 10, 13.5]$ is adjusted to $[0, 7.6, 10, 13.5]$, which is recorded as $[0, \gamma_1', \gamma_2', \gamma_3']$.

In order to separate the keywords with poor similarity in the density distribution, and further aggregate the keywords with similar attributes, so that the density distribution of words presents the aggregation effect of “synonyms attract each other”, which makes the identification of density class easier. This paper chooses the normal distribution function as the basis to define the weight ω, because the normal distribution converges at the values of zero and infinity, and there is a large rate of change near the standard deviation σ_w, which can play an activation role. When the distance between a pair of simple words is zero between two sets, the weight of the pair will not be infinite, leading to the replacement of the distance between the whole set, nor will it cause the weight of dissimilar words to be too small or even negative, ignoring the dissimilarity. At the same time, the significant change near the standard deviation can also play a nonlinear activation role. The weight $\omega_{1,i}'$ is defined as:

$$\omega_{1,i}' = \begin{cases} \left(e^{\frac{-(dis(v_{1,i}, cw_2))^2}{2\gamma_2'^2}} - e^{\frac{-(\gamma_3')^2}{2\gamma_2'^2}} \right) \Big/ 2 + e^{\frac{-(\gamma_3')^2}{2\gamma_2'^2}} & 0 < dis(v_{1,i}, cw_2) \le \gamma_3' \\ e^{\frac{-(\gamma_3')^2}{2\gamma_2'^2}} & \gamma_3' < dis(v_{1,i}, cw_2) \end{cases} \tag{9}$$

Substituting the experimental data of this paper based on the BERT vector model, the weight change curve is shown in **Fig. 3**. It can be seen from the results in **Fig. 3** that, compared with the standard normal distribution, the weight calculation proposed in this paper can achieve better separation in the middle of word meaning similarity, and better control the weight ratio between the same word pair and the word pair with completely different meaning.

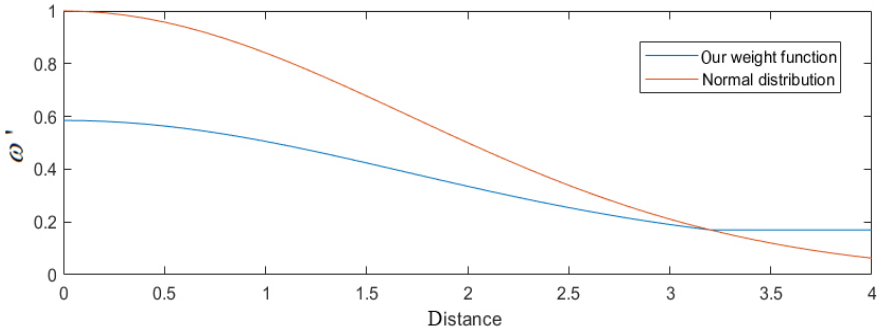

Fig. 3. Weight curve

For the multiple pairs of words in the set, the final weight is obtained after normalization as follows:

$$\omega_{1,i} = \omega'_{1,i} / (\sum_{i=1}^{k} \omega'_{1,i} + \sum_{j=1}^{k'} \omega'_{2,j}) \tag{10}$$

(2) Optimization of compound keyword density: In the optimization process of compound keyword density, first of all we should regard the word sense set of a compound keyword as a data point, in view of the nature of data density, the density effect between two data points is negatively related to the distance. Each data point has a density influence function, and the final density space is obtained by superposition of all data points. The density of compound keyword meaning set cw_i is defined as:

$$\rho_i = \sum_{j=1}^{n} e^{\frac{-Dis(cw_i, cw_j)^2}{2\sigma^2}} \tag{11}$$

In the above formula, $Dis(cw_i, cw_j)$ is the distance between two compound keyword sets; the standard deviation σ determines the size of normal distribution window. The value of standard deviation is to sort all distance data from small to large, and the distance in the top 2 ~ 10% is regarded as the standard deviation, which has good density recognition effect.

It should be noted that the nonlinear influence relationship is introduced in the distance measurement and density measurement of the set, but the connotation of the two measurement methods is not repeated. In the distance measurement, the adjustment of the weight of different information elements in the set is realized, which is the embodiment of the polarity of the internal matching degree. In the density measurement, the non-linearity in calculation is the adjustment of nonlinear density influence on the external distribution when the set is regarded as a point distributed in space after the distance is determined.

Clustering by Fast Search and Find of Density Peaks (CFSFDP) algorithm assigns two attributes to each data point: data point density ρ_i and nearest neighbor distance δ_i. In the process of calculation, if the density of data points is larger, the distance between the representative points is smaller. The distance that the density is larger than the point in the vicinity of the point is also smaller. However, if the point is a local maximum point in the

calculation process, it is necessary to cross the class to find a data point whose density is greater than the point, resulting that the nearest pointing distance is much greater than the normal distance value under the density. In this paper, we mainly use the nature of CFSFDP algorithm itself to set the threshold for qualitative division, and separate the points with larger density and distance as the class center. In the Formula 12, the nearest neighbor distance refers to the distance from the nearest point in all data points that are greater than the current point density, which can be defined as:

$$\delta_i = \begin{cases} \max\limits_{j} \left(Dis\left(cw_i, cw_j \right) \right) & \forall j, \rho_i > \rho_j \\ \min\limits_{j:\rho_j > \rho_i} \left(Dis\left(cw_i, cw_j \right) \right) & \text{others} \end{cases} \tag{12}$$

The CFSFDP algorithm points out that the greater the density of data points, the smaller the distance between the points, and the smaller the distance to the neighboring medium whose density is greater than the point. However, when the point is a local maximum point, it is necessary to find the data points whose density is greater than the point across classes, resulting in the nearest pointing distance far greater than the normal distance value at the density. The CFSFDP algorithm uses this property to qualitatively divide by artificially setting a threshold, and separates points with a larger density and a larger distance as the class center.

3.3 Class Label Generation

After clustering is completed, the keywords in the class need to be summarized to generate class labels. Taking the cluster1 keyword class as an example, the class label generation process is shown in **Fig. 4**.

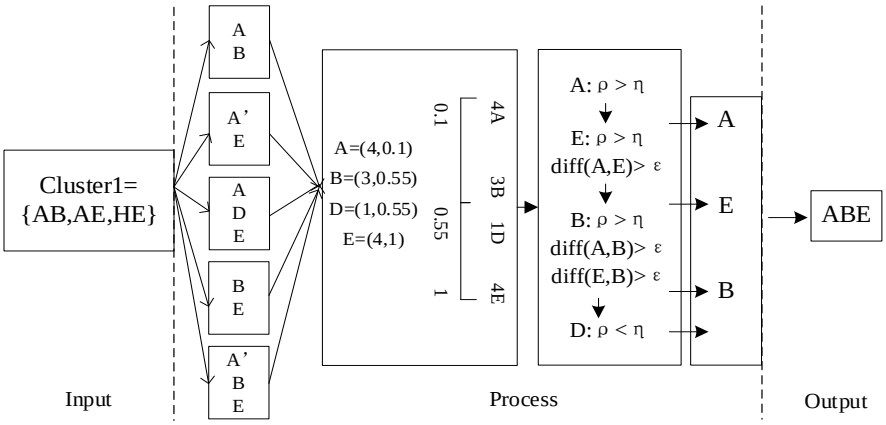

Fig. 4. Class label generation method

Step1: Count the word frequency and position order of simple words in the compound keyword set in the current class. The basis for judging whether simple words are synonyms is $dis'(v_{1,i}, v_{2,1}) < \gamma_1'$, when the position order is uniformly distributed, the mean value is taken; when the position order distribution is clustered, the most concentrated position is taken. Suppose a set of simple word order is $[x_1, x_2, \cdots, x_p]$, and the geometric ratio is mapped to the interval 0.1~1 as $[x_1', x_2', \cdots, x_p']$. Merging similar items in $[x_1', x_2', \cdots, x_p']$, starting from the order

with the most similar items, and merging from large to small.

Step2: Judging from the word frequency from large to small, the first word is the default output, and whether the i-th word is output is based on whether the word frequency is greater than the set threshold η, and whether the order difference $diff$ between the word frequency and the first $(i-1)$-th output words is greater than the set threshold ε.

Step3: Sort the output words according to the statistical rank value and combine them into class description labels.

4. Empirical Analysis

This study selects 21 management periodicals from CNKI(China National Knowledge Infrastructure) database as sample periodicals, the time of empirical data statistics is between 2017 and 2019, with a total of 14420 papers, and includes multiple types of datasets. The selected periodicals include Management World, Nankai Business Review, Journal of Public Management, Chinese Journal of Management Science, Management Science, Journal of Management Sciences in China, Chinese Journal of Management, Economic Management Journal, Journal of Industrial Engineering and Engineering Management, Systems Engineering-Theory Methodology Application, Operations Research and Management Science, Journal of Management, Modern Management Science, Scientific Decision-Making, Modernization of Management, Journal of Management Case Studies, Project Management Technology, Leadership Science, Shanghai Management Science, and Management Engineer. The keywords of each journal in every month in recent three years were counted.

4.1 Analysis of the Effect of Word Vector Acquisition

The input data set of the model constructed in this paper is the monthly keyword collection of all periodicals. The top 20 words in the monthly keyword popularity ranking are manually summarized and marked. The Synonyms are summarized in the marking process to conform to the actual statistics, which mainly include innovation and entrepreneurship, information transmission, green governance and organization. Taking the words of innovation and entrepreneurship as an example, in the process of obtaining the word vector, the keywords in the class need to be split in order. The keyword set obtained from the data is: innovation output, innovation diffusion, innovation ability, innovation investment, innovation output, innovation input, innovation incubation performance, and innovation cluster. The effect is shown in **Fig. 5**:

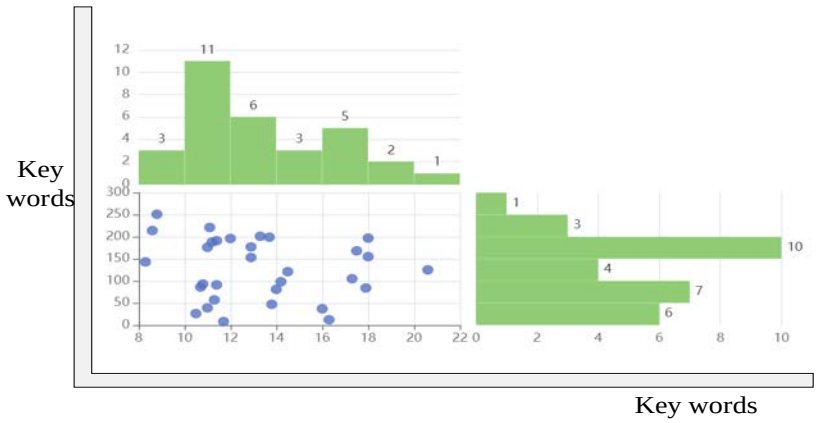

Fig. 5. Word vector keyword acquisition process fitting effect

It can be seen from **Fig. 5** that in the existing keywords, the fitting effect of most keywords is good. The influence of the innovation yield on the target is weak in the process of calculating the maximum correlation coefficient. The influence value is 1. Considering the accuracy of the model in the word sense distance measurement analysis and the hot keyword extraction accuracy analysis, the innovation yield in the keywords is removed. Retain the key words of innovation diffusion, innovation capability, innovation investment, innovation output, innovation input, innovation incubation performance and innovation cluster.

4.2 Semantic Distance Metric Analysis

Based on the semantic relationship, the keyword set is accurately mapped to the metric space to obtain an accurate distribution relationship, which is an important basis for the recognition of hot keywords. For the validity verification of word meaning measurement, the direct Bert model（i-BERT）distance measurement and the improved distance measurement are used to measure the labeled data respectively. Calculate the distance between the keyword category label and the keywords in the category, the sample results are shown in **Table 1**. Calculate the distance between hot keyword categories, the sample results are shown in **Table 2**.

Table 1. Distance between the keyword category label and the keywords in the category

Label	Method	Keyword						
innovation and entrepreneurship	-	innovation diffusion	innovation ability	innovation investment	innovation output	innovation input	innovation incubation performance	innovation cluster
	BERT	7.57	7.69	5.91	7.82	7.13	8.85	6.99
	i-BERT	6.37	6.50	4.81	6.62	5.94	7.65	5.79
information transfer	-	information sharing	information Gap	information integration	information support	information	information two-way	information

						interaction structure	transmission	dissemination
	BERT	8.17	7.71	6.62	7.79	8.04	5.27	4.11
	i-BERT	7.06	6.64	5.53	6.59	6.84	4.27	3.21
green governance	-	green administration	green governance mechanism	green governance subject	green transformation	green governance guidelines	green innovation strategy	-
	BERT	8.63	5.91	6.49	8.39	6.31	8.3	-
	i-BERT	7.43	4.81	5.29	6.91	5.03	7.10	-
organization	-	organizational virtue	organizational practices	organization structure	organizational innovation	project organization	organizational characteristics	-
	BERT	9.19	8.70	7.81	9.49	7.13	9.91	-
	i-BERT	8.00	7.50	6.61	8.77	5.91	8.71	-

Table 2. Distance between heat keywords

Label	BERT				i-BERT			
	innovation and entrepreneurship	information transfer	Green governance	organization	innovation and entrepreneurship	information transfer	green governance	organization
innovation and entrepreneurship	0	11.56	12.78	11.95	0	11.37	11.06	11.20
information transfer	11.56	0	11.96	11.27	11.37	0	12.27	10.75
green governance	12.78	11.96	0	14.02	11.06	12.27	0	12.23
organization	11.95	11.27	14.02	0	11.20	10.75	12.23	0

It can be seen from the experimental results in **Table 1** and **Table 2** that the distance measured by the proposed method is less than the vector Euclidean distance under the direct BERT model, but the distance shrinkage of the proposed method in the distance measurement between similar keywords is significantly greater than that between class labels. This is because the proposed distance measurement between composite keywords, on the one hand, will split the composite keywords into simple words for pairing, on the other hand, the adaptive weight distribution principle will automatically adjust the weight according to the distance between simple words, so that the weight of simple words with similar and the same semantics increases, so as to avoid the averaging of similar attributes. Among the class labels, the similarity between sub elements of different classes is low, and the weight adjustment effect of same-sex attraction is not significant.

4.3 Accuracy Analysis of Hot Keyword Extraction

Based on the analysis of word sense distance metric, the popularity keywords in words are extracted to further analyze the effectiveness of vector space mapping under the distance metric. Through the analysis of the existing research status, we can know that the commonly used methods for text feature extraction are TF-IDF and Word2vec. Therefore, the i-BERT model is compared with the original BERT model, the TF-IDF model, the Word2vec model and Decision Tree model. The paper selects the hot word recognition accuracy rate as the evaluation index, extracts the popularity of all journal keywords in the month, and determines the application effect of the model. The recognition process can be expressed by formula (13). The recognition process can be expressed by formula (13), which is:

$$AC = W / W\_ALL \tag{13}$$

where W denotes the number of hot words correctly identified in the current month, and W_ALL_i denotes the number of manually marked hot words in the current month.

The hot word recognition accuracy compares whether the final extracted hot keywords are correct. In order to further compare the accuracy of the word frequency statistics when extracting keywords, the recall rate is defined as:

$$RR = \frac{1}{n}\sum_{i=1}^{n} TP_i / T\_ALL_i \tag{14}$$

where TP denotes the correct count of the number of the current i-th hot word frequency, and T_ALL_i denotes the number of all samples under the hot word manually marked.

The data set in this study is divided into three sections, and the I-Bert model, BERT model, TF-IDF model, Word2vec model and decision tree model are used for three experimental verifications, and the mean of the experimental results is obtained. The accuracy results are shown in **Table 3**.

Table 3. Accuracy of keyword extraction

Data set	There are about 61% of the same words and 39% of the synonyms under the same generalization label in the labeled data set.				
Method	TFIDF	Word2vec	BERT	i-BERT	Decision Tree
Hot word recognition accuracy	66.4%	68.4%	74.4%	77.5%	66.3%
Recall rate	62.3%	67.5%	73.2%	82.1%	70.1%

From the experimental results in **Table 3**, it can be seen that the hot word recognition accuracy of i-BERT method can reach 77.5%, the recall rate can reach 82.1%, and the keyword extraction accuracy is better than the comparison algorithm. The keyword extraction based on TF-IDF method can not measure the similar meaning of words, resulting in a significantly lower recall and low precision of keyword extraction. The keyword extraction based on word2vec method and BERT method is based on word vector, which can measure the meaning of words and make recall rate and accuracy higher. However, word2vec method is highly dependent on word segmentation, which leads to the inability to flexibly express compound keywords. Compared with word2vec method, BERT method can solve the problem of polysemous words properly. Because there are few polysemous words in the keywords of the paper even after word segmentation, the advantage of polysemous words is not significant, but it can express the vector of any compound keywords, which makes the vectorization more comprehensive, and improves the accuracy and recall rate again.

5. Conclusion

In order to improve the accuracy of subject hot keyword acquisition, this paper proposes a hot keyword extraction method of sci-tech periodicals based on improved BERT (i-BERT), which is used to accurately capture the hot issues of subject research and provide effective support for journal planning. The conclusions are as follows:

(1) Make full use of the BERT language model to dynamically generate the context semantic representation of characters through the two-way transformer structure, which can better represent the semantic and sentence characteristics of characters than the traditional word embedded vector representation. The distance of composite keyword set is defined innovatively, and the distance weight of each pair of matching elements in the set adopts normalized nonlinear mapping to overcome the disadvantage of bad correspondence between distance value.

(2) The data set selected in the experiment is the keywords given by the literature. The data set can be the abstract of the literature or even the full text content by improving the model so as to improve the accuracy of the discovery of research hot areas or directions through the comprehensiveness of the data set.

(3) The extraction method of subject hotspots proposed in this paper, which is based on the literature keywords in a relatively new period of time, can further strengthen the accuracy and effectiveness of timeliness. Future researches can study the evolution law of subject hotspots from the literature data of a larger time span, and study the hot words and new words from the literature data of a larger subject field or even across disciplines. Future researches can also predict the changing trend of discipline hot spots that enables the sci-tech periodicals to guide the direction of discuripline research.

(4) The empirical analysis shows that the extraction accuracy of hot keywords of sci-tech periodicals proposed in this paper reaches 77% and the recall rate reaches 82%, which is higher than the existing common journal keyword extraction methods, and can ensure the accuracy and timeliness of capturing hot issues in discipline research.

Acknowledgement

This work was financially supported by the National Natural Science Foundation of China (71874019).

References

[1] Qiao, Wenchuan, Zheng Fang, and Bailu Si, "A sampling-based multi-tree fusion algorithm for frontier detection," *International Journal of Advanced Robotic Systems*, vol. 16, no. 4, pp. 1-14, 2019. Article (CrossRef Link)

[2] Liu, Bing, Zhijun Lv, Nan Zhu, and Dongyu Chang, "Research on the evaluation of the dissemination ability of Sci-Tech periodicals based on hesitant fuzzy linguistic," *International Journal of Uncertainty, Fuzziness and Knowledge-Based System*, vol. s 28, no.02, pp.153-167, 2020. Article (CrossRef Link)

[3] Anzoise, Valentina, Debora Slanzi, and Irene Poli, "Local stakeholders' narratives about large-scale urban development: The Zhejiang Hangzhou future Sci-Tech City," *Urban Studies*, vol. 57, no. 3, pp.655-671, 2020. Article (CrossRef Link)

[4] Jin, Yuran, and Xin Li, "Visualizing the hotspots and emerging trends of multimedia big data through scientometrics," *Multimedia Tools and Applications*, vol. 78, no. 2, pp. 1289-1313, 2019. Article (CrossRef Link)

[5] Bin, Cheng, Chen Weiqi, Chu Shaoling, and Hu Chunxia, "Visual Analysis of Research Hot Spots, Characteristics, and Dynamic Evolution of International Competitive Basketball Based on Knowledge Mapping," *SAGE Open*, vol. 11, no. 1, pp.1-13, 2021. Article (CrossRef Link)

[6] Zhao, Jian, Guanyu Yu, Mengxi Cai, Xiao Lei, Yanyong Yang, Qijin Wang, and Xiao Zhai, "Bibliometric analysis of global scientific activity on umbilical cord mesenchymal stem cells: a swiftly expanding and shifting focus," *Stem cell research & therapy*, vol. 9, no. 1, pp. 1-9, 2018. Article (CrossRef Link)

[7] Yu, Ma, Su Shuang, and Liu Jie, "Knowledge map of career research in China over past 20 years-CiteSpace bibliometric analysis based on CSSCI journals," in *Proc. of the 2019 Annual Meeting on Management Engineering*, Kuala Lumpur, Malaysia, pp. 154-159, 2019. Article (CrossRef Link)

[8] Zhao, Jian, Guanyu Yu, Mengxi Cai, Xiao Lei, Yanyong Yang, Qijin Wang, and Xiao Zhai, "Bibliometric analysis of global scientific activity on umbilical cord mesenchymal stem cells: a swiftly expanding and shifting focus," *Stem cell research & therapy*, vol. 9, no. 1, pp.1-9, 2018. Article (CrossRef Link).

[9] Feng, Zhong-kai, Wen-jing Niu, Zheng-yang Tang, Zhi-qiang Jiang, Yang Xu, Yi Liu, and Hai-rong Zhang, "Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization," *Journal of Hydrology*, vol. 583, pp.124-135, 2020. Article (CrossRef Link).
[10] Huertas-Valdivia, Irene, Anna Maria Ferrari, Davide Settembre-Blundo, and Fernando E. García-Muiña, "Social life-cycle assessment: A review by bibliometric analysis," *Sustainability*, vol. 12, no. 15, pp.62-73, 2020. Article (CrossRef Link)
[11] Ding, Yi, and Xian Fu, "The research of text mining based on self-organizing maps," *Procedia Engineering*, vol. 29, pp. 537-541, 2012. Article (CrossRef Link)
[12] Wang, Bojie, Qin Zhang, and Fengqi Cui, "Scientific research on ecosystem services and human well-being: A bibliometric analysis," *Ecological Indicators*, vol. 125, pp.107-115, 2021. Article (CrossRef Link)
[13] Lei, Hongzhen, and Xiaoli Chen, "Hot Spots and Trends of Spillover Effects of Brand Scandals: Visual Analysis Based on Citespace," in *Proc. of 2020 International Conference on Modern Education and Information Management (ICMEIM)*, Dalian, China, pp.607-610, 25-27 Sept, 2020. Article (CrossRef Link)
[14] Tang, Zhong, Wenqiang Li, Yan Li, Wu Zhao, and Song Li, "Several alternative term weighting methods for text representation and classification," *Knowledge-Based Systems*, vol. 207, pp. 109-121, 2020. Article (CrossRef Link)
[15] Mingxi Zhang, Xuemin Li, Shuibo Yue, and Liuqian Yang, "An empirical study of TextRank for keyword extraction," *IEEE Access*, vol. 8, pp. 178849-178858, 2020. Article (CrossRef Link)
[16] Bhatta, Janardan, Dipesh Shrestha, Santosh Nepal, Saurav Pandey, and Shekhar Koirala, "Efficient estimation of Nepali word representations in vector space," *Journal of Innovations in Engineering Education*, vol. 3, no. 1, pp. 71-77, 2020. Article (CrossRef Link)
[17] Chen, Ziyan, Yu Huang, Yuexian Liang, Yang Wang, Xingyu Fu, and Kun Fu, "RGloVe: An improved approach of global vectors for distributional entity relation representation," *Algorithms*, vol. 10, no. 2, pp. 42-53, 2017. Article (CrossRef Link)
[18] Sarzynska-Wawer, Justyna, Aleksander Wawer, Aleksandra Pawlak, Julia Szymanowska, Izabela Stefaniak, Michal Jarkiewicz, and Lukasz Okruszek, "Detecting formal thought disorder by deep contextualized word representations," *Psychiatry Research*, vol. 304, pp.114-135, 2015. Article (CrossRef Link)
[19] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *Computation and Language*, pp.4171-4186, 2018. Article (CrossRef Link)
[20] Mehmood, Rashid, Guangzhi Zhang, Rongfang Bie, Hassan Dawood, and Haseeb Ahmad, "Clustering by fast search and find of density peaks via heat diffusion," *Neurocomputing*, vol. 208, pp. 210-217, 2016. Article (CrossRef Link)
[21] Mohammadi Ehsan, and Karami Amir, "Exploring research trends in big data across disciplines: A text mining analysis," *Journal of Information Science*, vol. 48, no. 1, pp. 44-56, 2022. Article (CrossRef Link)
[22] Bayrak Tuncay, "A comparative analysis of the world's constitutions: a text mining approach". *Social Network Analysis and Mining*, vol. 12, no. 1, pp. 00857, 2022. Article (CrossRef Link)
[23] Zhang Yali, and Shang Pengjian, "KM-MIC: An improved maximum information coefficient based on K-Medoids clustering," *Communications in Nonlinear Science and Numerical Simulation*, vol. 111, no. 8, pp. 106418, 2022. Article (CrossRef Link)

Bing Liu is currently pursuing a Doctoral Degree in Management at School of Economics and Management of Dalian University of Technology. His research focuses on fuzzy logic and journal evaluation.

Zhijun Lv received his Ph.D. degree in Management from Dalian University of Technology in 2010. He is a Professor of Editorship at Dalian University of Technology Press. His research focuses on enterprise management and culture industry management.

Nan Zhu is currently pursuing a Doctoral Degree in Management at School of Economics and Management of Dalian University of Technology. His research focuses on enterprise management.

Dongyu Chang is currently studying for a Master's Degree in Management at School of Economics and Management of Dalian University of Technology. Her research focuses on operations management and journal internationalization.

Mengxin Lu is currently studying for a Master's Degree in Management at School of Economics and Management of Dalian University of Technology. Her research focuses on operations management and journal user research.