American Journal of Computer Science and Technology
2022; 5(2): 134-145
http://www.sciencepublishinggroup.com/j/ajcst
doi: 10.11648/j.ajcst.20220502.25
ISSN: 2640-0111 (Print); ISSN: 2640-012X (Online)

Design and Implementation of Journal Recommendation Model Based on L-BERT

Peng Liu[1, 2, 3], Yuzhi Xiao[1, 2, 3, *], Youpeng Qin[1, 2, 3], Lin Zhou[1, 2, 3]

[1]School of Computer, Qinghai Normal University, Xining, China

[2]Tibetan Information Processing and Machine Translation Key Laboratory of Qinghai Province, Qinghai Normal University, Xining, China

[3]Key Laboratory of Tibetan Information Processing, Qinghai Normal University, Xining, China

Email address:

1531453132@qq.com (Peng Liu), 56188111@qq.com (Yuzhi Xiao), 853613572@qq.com (Youpeng Qin), 2245774955@qq.com (Lin Zhou)

[*]Corresponding author

To cite this article:

Peng Liu, Yuzhi Xiao, Youpeng Qin, Lin Zhou. Design and Implementation of Journal Recommendation Model Based on L-BERT. *American Journal of Computer Science and Technology*. Vol. 5, No. 2, 2022, pp. 134-145. doi: 10.11648/j.ajcst.20220502.25

Received: April 13, 2022; **Accepted**: May 14, 2022; **Published**: June 9, 2022

Abstract: The output cycle and quantity of academic achievements of researchers can promote academic exchanges and integration to a certain extent. Therefore, it is particularly important to design a journal recommendation model for scientific researchers that is fast and relatively suitable for producing scientific research results. Based on this, focusing on the information knowledge of papers and the characteristics of journal topics, combined with deep learning technology, this paper proposes a journal-oriented recommendation model based on L-BERT (LDA-BERT), which realizes the purpose of recommending journals for scholars. Firstly, topic extraction is performed on papers in a specific journal, and the number of topic words is determined by perplexity. Secondly, the vectorized representation of the topic words is carried out by using the BERT model, and the fixed representation value is obtained through the mean value idea. Finally, Euclidean distance is used to determine the similarity between papers and journals. In order to verify the effect of L-BERT model, it is compared with Word2Vec model on the real data set. The results show that the recommended results of L-BERT model are improved by 17%, 10% and 13% respectively in Precision, Recall, and F1-Score, which fully shows that this research has a certain application value, and the research results can help researchers shorten the cycle of looking for journals.

Keywords: LDA Model, BERT Model, Euclidean Distance

1. Introduction

In recent years, with the further development of the Internet, the speed of information and data generation and dissemination has become faster and faster. The data available to people has become more and more diverse in fields, more and more complex in form, and more and more grand in scale, people are entering the “Era of Big Data” [1, 2]. The continuous development of Big Data promotes the innovation and progress of science and technology, prompting scientific researchers in all walks of life to continuously generate massive amounts of data when exploring knowledge, and the academic field has ushered in the research direction of “Academic Big Data”. Academic Big Data covers a wealth of academic research information, including academic journals, dissertations, patents and other academic-related data. Similar to the characteristics of Big Data, Academic Big Data also has the characteristics of Volume, Variety, Value, Veracity, and Velocity [3].

The development of Academic Big Data not only helps scholars better grasp academic information and promotes academic efficiency, but also brings challenges that traditional statistical analysis data methods cannot meet the functional processing of academic big data such as storage and analysis. In this context, many publications related to different professional fields have emerged, For example, The core journals of computer include “Computer Application Research”, “Computer Engineering”, etc. The core journals of normal education include “People's Education”, “Comparative Education Research”, etc. When researchers face a large number of journal publishing institutions, if they submit scientific achievements to inappropriate publishing institutions, it is very likely that the paper will be rejected, which is not conducive to the publication of scholars' achievements and reduces their scientific research work. effectiveness. Therefore, how to effectively recommend the publication site to scholars has become a key issue. To deal with these problems, Wang et al. extracted features from the paper data through information gain technology, and combined with the content-based recommendation algorithm, proposed a journal recommendation algorithm. Wang introduced topic distribution technology, combined with the publication time of the paper, and recommended for scholars the relevant topics with time factor to belong to the paper publishing institution. Yang et al. by analyzing the style of scholars' submitted papers, a recommendation model based on collaborative filtering strategy is proposed. However, most of the studies are based on traditional recommendation algorithms, and there are still gaps and deficiencies in the research on journal recommendation which combined with topic models in specific fields [4-6].

Based on this, by the Scrapy distributed crawler tool, this paper first obtains computer-related academic data from CNKI, that including 36 journals in the past ten years (2010/1/1-2020/12/31). Secondly, based on the word segmentation technology and the LDA (Latent Dirichlet Allocation) topic model, the experimental data is clustered by topic, and the "journal-keywords" data model is constructed combined with the perplexity. Third, use the BERT (Bidirectional Encoder Representations from Transformer) model to perform word embedding processing on the processed keywords, so that all of them are converted into vector form. Finally, the similarity between the input item and each recommended journal is measured by the Euclidean distance value, and based on this, the corresponding journals are recommended for scholars. At the same time, the experimental results were analyzed with Precision, Recall, and F1-Score as the evaluation indicators of the recommended effect.

2. Related Theory

2.1. LDA Model

LDA is a generative topic model for unsupervised learning [7]. It is considered that each word in an article is generated through the process of "the article selects a certain topic through a certain probability, and the topic selects a certain vocabulary through a certain probability", "document-topic" and "topic-word" are subject to polynomial probability distribution. It is also considered a three-layer Bayesian probability model [8], it contains a three-layer structure of document-topic-word, and its topology is shown in Figure 1:

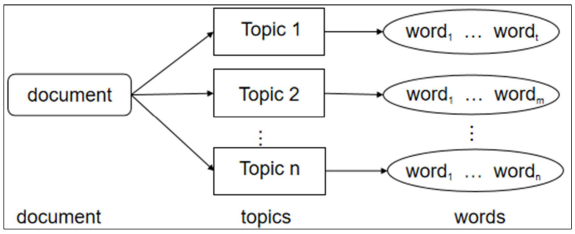

Figure 1. *LDA three-layer topology.*

In large-scale data texts, the application of the LDA model is helpful for the mining of potential topic information in the text, which can find the focus of a specific text dataset, and better perform text modeling.

The LDA model is shown in Figure 2:

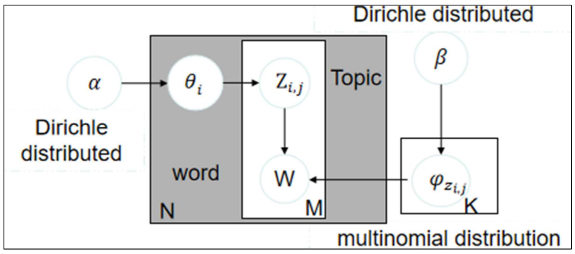

Figure 2. *LDA model.*

From the LDA model principle, it can be deduced that the joint probability distribution of topic and word in the generation process can be expressed as the following formula:

$$p(W|d_i)=\sum_{K=1}^{K} p\left(W\middle|Z_{i,j}\right)p(Z_{i,j}|d_i) \tag{1}$$

The meanings of the LDA model and related joint probability mathematical symbols are shown in Table 1:

Table 1. *LDA Model and Joint Probability Mathematical Notation.*

Notation	Meaning
φ	topic correspondence vocabulary multinomial distribution
θ	multinomial distribution of documents corresponding to topics
K	total number of topics
d_i	document i
$Z_{i,i}$	the topic of the j-th word in document i
N	documentation set
W	generate vocabulary
M	all vocabulary contained in this document
Α	prior parameters for the probability distribution of document-topic
β	prior parameters for the probability distribution of topic-word

The document generation process of the LDA model is described in Table 2:

Table 2. *LDA document generation process.*

Document Generation Step s for LDA Topic Models	
Step 1	Based on Dirichlet distribution α and sampling algorithm, obtain topic distribution θ_i in document i, $\theta_i \sim Dir(\alpha)$.
Step 2	For word j in document i, identify a potential topic $Z_{i,j}$ from θ_i for it, $Z_{i,j} \sim Mult(\theta_i)$.
Step 3	Based on the Dirichlet distribution β and the sampling algorithm, the lexical polynomial distribution $\varphi_{Z_{i,j}}$ corresponding to topic $Z_{i,j}$ is obtained, $\varphi_{Z_{i,j}} \sim Dir(\beta)$, $k \in [1, K]$.
Step 4	Consider topic $Z_{i,j}$ in Step 2 and multi-nomial distribution $\varphi_{Z_{i,j}}$ in Step 3, extract a topic word $W_{i,j}$.
Step 5	Iteratively execute Step 2, Step 3, and Step 4 until all vocabulary extraction is completed.

2.2. BERT Model

BERT (Bidirectional Encoder Representations from Transformers) [9] is a pre-trained language model released by Google in 2018. It runs based on the Encoder mechanism of Transformer [10], considering the context of the term, the corpus is trained in both directions. When the corpus is sufficient, the word vector obtained by training has strong representation and fit, which improves the problem of semantic aggregation in the word embedding training of Word2Vec [11]. It effectively solves the problem of difficulty in distinguishing word meanings in traditional training models.

2.2.1. The Overall Architecture of the BERT Model

The BERT structure consists of Input Layer, Encoder Layer, and Output Layer, as follows:

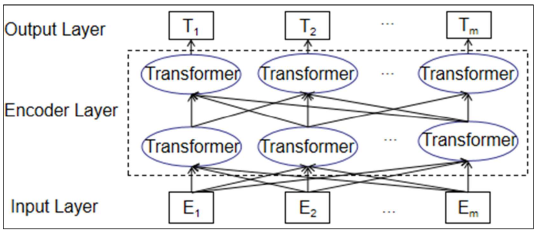

Figure 3. *The model architecture of BERT.*

E_1, E_2,..., E_m are input items, the intermediate encoding layer is composed of multiple Transformer-based Encoder structures stacked, T_1, T_2,..., T_m are the output items of the model.

2.2.2. Input to the BERT Model

The BERT model processes the corpus data in units of token, the input vector of the model is the result vector after the summation of three vectors: Token Embedding, Segment Embedding, and Position Embedding. Take “All right? Yes” as an example, as follows:

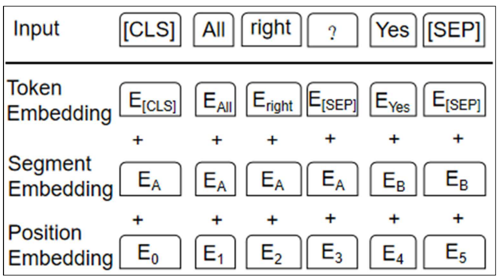

Figure 4. *BERT model input Item.*

Token Embedding is the word embedding vector. Position Embedding is the position vector of the word. Segment Embedding is the sentence vector. The words in the previous sentence are represented by EA, and the words in the latter sentence are represented by EB. The symbol [CLS] is embedded before the first character of the input sentence, representing the start of a sentence. The symbol [SEP] is used to separate the two input sentences so that the BERT model can classify and predict the sentences.

2.2.3. Transformer Encoder

The core component of the BERT model is the Transformer's Encoder [12], each Encoder is composed of a Self-Attention mechanism and a Feed Forward neural network. The architecture of Encoder is shown in Figure 5:

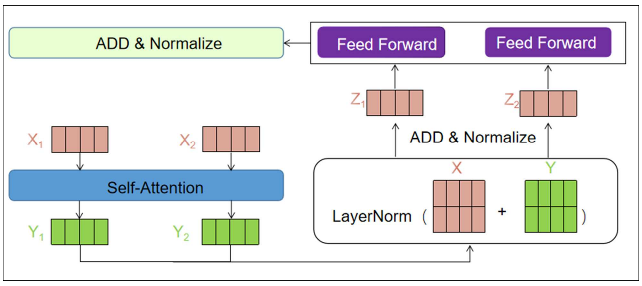

Figure 5. *The architecture of Encoder.*

Assume that the two result vectors obtained by the summation method are X_1 and X_2, respectively. After inputting it into the Encoder, the input vector is processed by the Self-Attention layer to obtain the vectors Y_1 and Y_2. The processing process is as follows:

(1) First, in order to effectively use the context information in the calculation process, the matrices Q, K, V are obtained by linearly transforming X, and the transformation formula is as follows:

$$Q = XW^Q \tag{2}$$

$$K = XW^K \tag{3}$$

$$V = XW^V \tag{4}$$

W^Q, W^K, W^V are the randomly initialized matrix.

(2) Secondly, score the contribution of other words to the current word by performing dot product of Q and K^T, softmax normalize the dot product result after dividing it by the value of a constant $\sqrt{d_k}$, Multiply the normalized value as a weight with the initial matrix V to obtain a new word vector matrix, the computational mathematical expression of Self-Attention is as follows:

$$\text{Attention}(Q,K,V)=\text{Softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{5}$$

$\sqrt{d_k}$ is the square root of the value that first dimension of the initialization matrix, $QK^T/\sqrt{d_k}$ is to make the attention matrix have a standard normal distribution.

(3) Finally, BERT introduces a Multi-head Attention mechanism, assuming that the number of heads is n, n times of Self-Attention calculations will be performed to obtain n times new word vector matrices, all matrices will be spliced horizontally to form a new matrix G, and then a matrix WO that can be multiplied by G will be randomly initialized, G*WO gets the vector matrix Y of the vectors Y_1 and Y_2.

(4) After getting the vector matrix Y, Norm&Normalize [13] it with the original vector matrix X, get two new vectors Z_1 and Z_2, pass it to the Feed Forward neural network, and the Feed Forward neural network will use the final vector obtained as the input vector of the next Encoder after residual connection and normalization.

2.3. Similarity Index

The method of the recommendation system [14] is usually to map the experimental text data into the vector space through model training, to obtain the vector representation of the corresponding words. The similarity between the vector and other text vectors is measured according to the size of the calculated value of the similarity index. The larger the similarity value, the more similar they are. Commonly used similarity calculation indicators are Euclidean distance and cosine similarity.

(1) Euclidean distance

Euclidean distance [15] is based on the absolute distance between points in the space as an index to measure the degree of similarity, and its numerical value is directly related to the coordinate position of the point in the space. It is usually used to measure the similarity of a single individual integrated in multiple dimensions. The mathematical expression for calculating the Euclidean distance between d_1 and d_2 is as follows:

$$\text{Euclidean}(d_1,d_2)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \tag{6}$$

Among them, n represents the dimension of individual x and individual y.

(2) Cosine similarity

Cosine similarity [16] is one of the methods used to measure the degree of similarity between documents in the process of text data processing. It is based on the idea of cosine calculation in mathematical operations. Assuming that word frequency is used as the vector dimension value, the mathematical expression for calculating the cosine similarity value between document d_1 and document d_2 is as follows:

$$\text{Cosine}(d_1,d_2)=\frac{\sum_{i=1}^{n}(X_i\times Y_i)}{\sqrt{\sum_{i=1}^{n}(X_i)^2\times\sum_{i=1}^{n}(Y_i)^2}} \tag{7}$$

Among them, n represents the total number of different words after word segmentation, X_i represents the word frequency of the i-th word in the document d_1, and Y_i represents the word frequency of the i-th word in the document d_2.

3. Experimental Basis and Dataset

3.1. Experimental Basis

When conducting researches related to recommendation systems, the support of computer software and hardware is often required. The configuration of the experimental operating environment in this paper is shown in Table 3:

Table 3. *Environment for journal recommendation experiment.*

Experimental environment configuration
CPU: Intel second generation Core i7
System Type: 64 bit operating system, Memory: 8G
Operating System: Microsoft windows10 professional
Development Environment: Python+Pycharm2020 Professional Edition
Database: MySQL+Navicat for MySQL Enterprise Edition
Crawler Tool: Scrapy distributed crawler system

3.2. Experimental Dataset

The raw data of the experiment in this paper is the computer related academic data obtained from CNKI based on the Scrapy distributed crawler tool. The data acquisition process is shown in Figure 6:

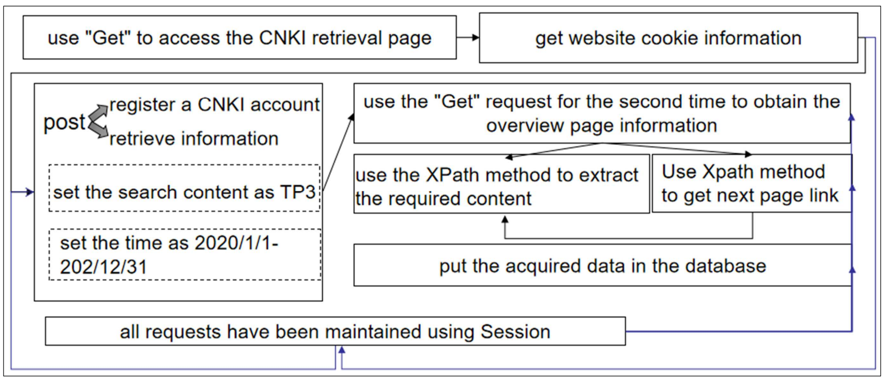

Figure 6. *Data acquisition process.*

First, determine the target website to collect data, this paper uses CNKI as the data source to obtain the website. Secondly, after entering the website, obtain the cookie information of the current web page, and use it as the cookie value for all web page operations. Again, make a POST request to the current web page, the information class is set to TP3, and the time is set to 2010/1/1-2020/12/31. Then, according to the jumped web page information, based on the Xpath technology, the data value contained in the current page and the web page link of the next page are analyzed, and the web page information is mined circularly until the ideal scale dataset is obtained. Finally, the acquired data is stored in a one-to-one correspondence between fields and database tables to prepare for subsequent experiments.

In order to achieve the effect of journal recommendation more accurately, the 36 types of journals with the largest number of papers will be intercepted from the obtained dataset as the research objects. Limit the experimental dataset to: Data records of 21,700 papers published in 36 journals with computer-related fields within 10 years (2010/1/1-2020/12/ 31) obtained from CNKI.

author	orgn	abstract	keywords	journal	publishTim	cite	down	classifyNu
Chen Jiyi; he Tao; Hu Jie	Shanghai Jiao Tong	A regulariza	CONFOCAL; Image R	Computer system a	2020-02-1	0	60	TP391.41
Wang Dongbin;Hu Mingzen	Research Center for	The operati	topology informatio	computer engineeri	2010-03-2	0	73	TP393.06
Hao Liang;Cui Gang;Qu Min	School of Computer	In view of th	cloud computing; re	Intelligent Compute	2014-10-0	4	190	TP3
Li Zhiyu; Lin Jiarui; Sun Yanb	State Key Laborator	The measur	vision measurement;	Acta Optics	2020-05-1	0	133	TG83;TP2
Wang Xingjian;Chen Ping;Ti	Beijing Normal Unive	In order to	Unified Communicat	Educational Informa	2020-02-0	0	35	G647;TP3
Wang Xingjia;Dong Lina;Li C	Department of Elect	A new type	left ventricle; segme	Chinese Journal of B	2011-08-2	7	88	TP391.41
Huang Zhiwen	Beijing Institute of T	The rapid d	computer technolog	China New Commur	2017-02-0	1	68	TP3
Jiang Xuesong; Yao Hongxu	School of Computer	As an impor	night image enhance	Intelligent Compute	2020-03-0	0	206	TP391.41
Kang Shanshan; Tang Xiaoq	Department of Mec	Aiming at th	motor drive platforn	Experimental Techn	2017-09-1	7	221	TH122;TP
Xu Xue	School of Informatio	With the ac	data security; data p	China Science and T	2012-08-1	4	126	TP309
Chen Gang; Yan Yingzhan; Li	Department of Com	The field of	association rules; fre	Microelectronics an	2014-01-0	5	81	TP311.13
Zhang Peiying;Fang Longyu	School of Computer	Word simila	Word similarity; How	computer technolo	2014-09-1	8	192	TP391.1
Wang Jie; Yu Yanshuo; Zhou	Department of Embe	Web tags he	Web label clustering	computer science	2014-12-1	5	100	TP311.13
Yang Qiuhui;Hong Mei;Guo	School of Software,	The practica	software testing; pra	computer educatio	2016-02-1	15	188	G642;TP3

Figure 7. *Part of the experimental dataset.*

Figure 7 is the storage status of some of the obtained dataset in the database.

4. Experimental and Result Analysis

4.1. Determination of Topic Categories

In the LDA model, the determination of the number of potential topics K needs to be set manually. The value of K directly affects the training effect. The selection of the appropriate value of K has become an important exploration problem. In general, use Perplexity [17] as a determinant indicator of the K value. It can judge the quality of the model results according to the number of topics in the LDA model within a set range. The lower the perplexity, the more obvious the clustering, and the better the training effect of the model. In order to obtain the appropriate number of potential topics, uses the principle of perplexity to determine the number of topics in the journal. To prevent the occurrence of one topic, the value of the number of topics in the calculation of perplexity is set between 2 and 30. Topic keywords are set to 5. The calculation formula of perplexity is shown in (8):

$$\text{Perplexity}=\exp\left\{\frac{-\sum_{d=1}^{N}\log(p(w_d))}{\sum_{d=1}^{N}T_d}\right\} \tag{8}$$

N represents the total number of journal document sets, T_d represents the number of words in the d-th paper, and $p(w_d)$ refers to the probability of each word appearing in document d. The mathematical expression of $p(w_d)$ is shown in (9):

$$p(w_d)=\sum_z p(w|z)p(z|d) \tag{9}$$

z is the topic, d is the document, and w is the vocabulary.

Based on the experimental dataset, the "topic-perplexity" distribution of the topic number between [2, 30] and the keyword information of the corresponding topic were obtained for 36 journals respectively. Taking two journals of Computer Engineering and Design and Computer Science as examples, the distribution of "topic-perplexity" and "topic-keywords" is shown in Figure 8 and Figure 9:

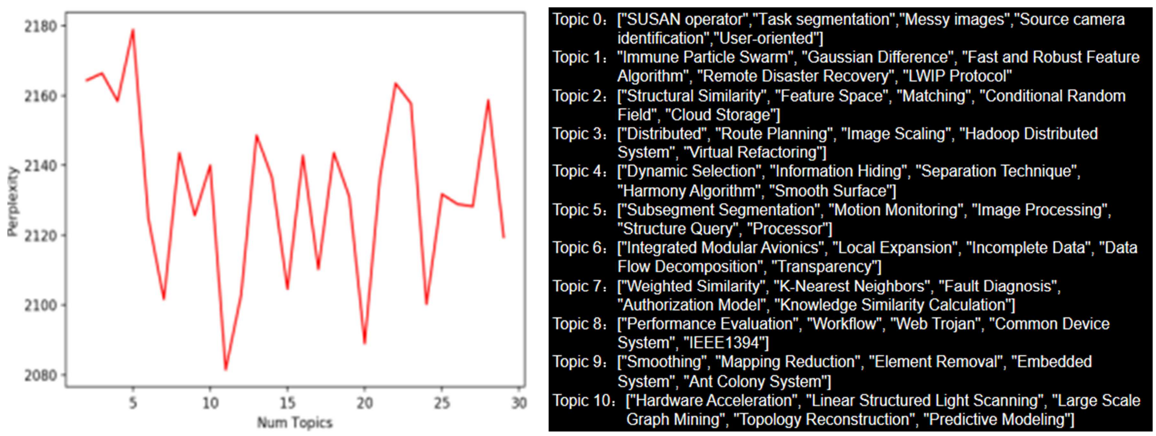

Figure 8. *Distribution map of Computer Engineering and Design.*

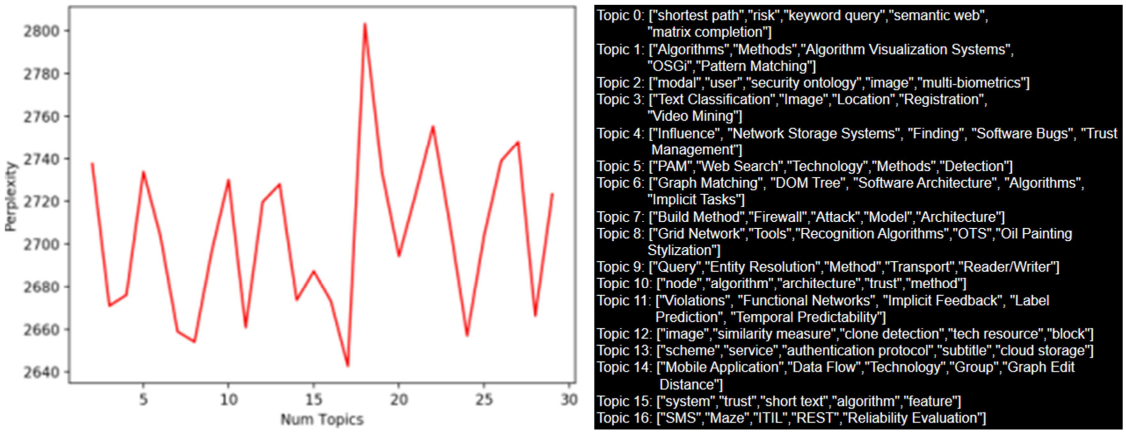

Figure 9. *Distribution map of Computer Science.*

The "topic-perplexity" of 36 kinds of journals was calculated one by one, and the statistics were carried out in the form of initial letter combination of journal names (CE means Computer Engineering, CEdu means Computer Education. CS means Computer Simulation, CSci means Computer Science), the number of papers in each type of journal and the distribution of the number of topics with the least confusion are shown in the figure below:

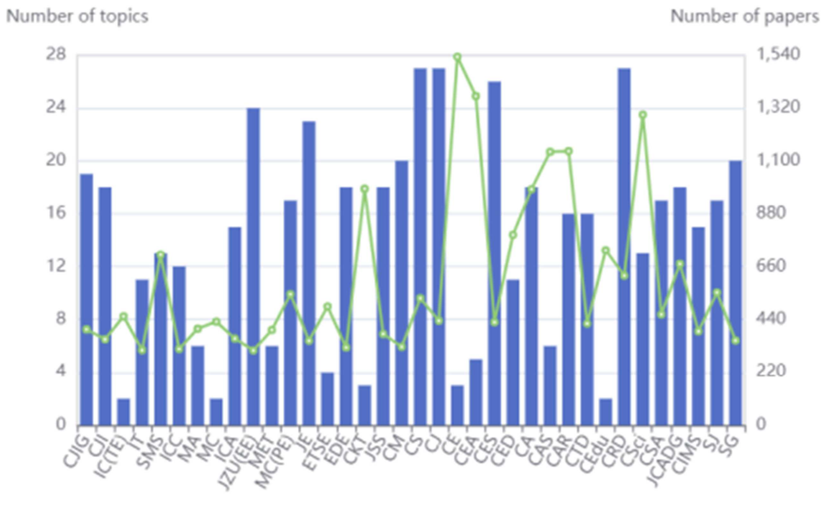

Figure 10. *Distribution of journal papers and topics.*

4.2. Journal Recommendation Based on L-BERT Model

In order to improve the accuracy of journal recommendation, combined with the attributes of paper portrait and journal portrait, data models such as "Journal-Papers", "Journal-keywords" and "Topic- keywords" are constructed. In the process of model establishment, aiming at the problems of data diversity, high dimension and sparsity of data relations, a new recommendation model which integrating LDA topic model and BERT training language model is proposed. It is called LDA-BERT model, abbreviated as L-BERT model. The implementation Step s of L-BERT model are shown in Table 4:

Table 4. *L-BERT model implementation Step s.*

Model: L-BERT	
Step 1	Jieba word segmentation for the input content
Step 2	According to the user-defined library, the word segmentation results are processed to stop words, and then the processing results are stored in the document
Step 3	The LDA model models the results in Step 2
Step 4	The modeling result is combined with perplexity index to determine the value of topic K, the "Journal-Keywords" data model is obtained
Step 5	Perform word embedding processing on the "Journal-Keywords" data model through the BERT model
Step 6	Store the keywords' vector and represent the journal vector as the average sum of all keyword vectors in the journal

Journal recommendation is a process of data combination and model operation. First, obtain the experimental dataset and take the data preprocessing. Secondly, based on the processed data, the fields for constructing the journal portrait are summarized by considering the journal name, journal coverage information and other attributes. Thirdly, according to the processing results, the dataset is segmented, perplexed, and divided into topics. Then, the results are processed by L-BERT model to embed related words to construct journal portraits. Finally, after the input information is also processed by the L-BERT model, the Euclidean distance is solved with the journal vector, calculate the similarity between them and make recommendations based on this.

After processing the experimental data by the L-BERT model, each journal contains 5K keywords, suppose the keywords of journal i are i_1, i_2,..., i_{5k}, then the vector corresponding to each keyword is $\overrightarrow{i_1}, \overrightarrow{i_2}, \ldots, \overrightarrow{i_{5K}}$. The technical roadmap for journal recommendation based on the L-BERT model is as follows:

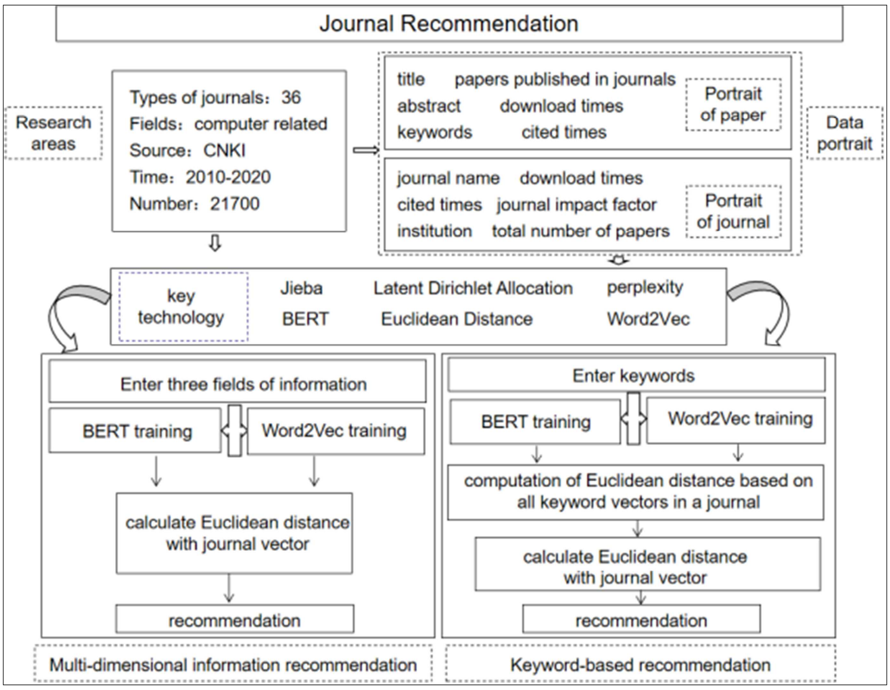

Figure 11. *Journal recommendation technology roadmap.*

Journal recommendation is divided into two modules: Multi-dimensional information recommendation and Keyword-based recommendation. Multi-dimensional information recommendation refers to three input fields, they are the title, abstract and keywords of the paper to be recommended (the multi-keywords are separated by spaces). Keyword-based recommendation can be subdivided into two types: One is recommendation for the similarity between multi-keywords and journals, the other is recommendation for the similarity between multi-keywords and all keywords that in journals. Among them, the journal vector is represented by the average value of all keywords vectors in the journal, that is $\sum_{j=1}^{5K} i_d/5K$.

(1) Recommendation for the similarity between multi-keywords and journals. Firstly, the input keywords are transformed into input vectors by using L-BERT model, and if there are multi-keywords, find the vector average of multi-keywords as the input vector. Second, find the Euclidean distance between the input vector and the journal vector. Finally, the corresponding journals are displayed as the recommended results in descending order of similarity. Taking "complex network" and "social computing" as input keywords as examples, the results of journal recommendation are shown in Figure 12:

```
Loading model cost 1.241 seconds .
Prefix dict has been built successfully .
Please enter key words and separated by spaces: complex network social computing
[('Computer and Modernization','4.704773426055908'), ('Software Guide','4.732361137866974'),
('Journal of Computer Research and Development','4.806760907173157'), ('Computer Systems
Applications','4.845077753067017'),('Intelligent Computer and Applications','4.864785552024841'),
('Computer Science and Engineering','4.932655215263367'), ('Electronic Design Engineering','5.00
1298904418945'), ('Chinese Journal of Computers','5.04720938205719'), ('Journal of Chinese
Information Processing','5.080924987792969'), ('Journal of Zhejiang University-Science A','5.08762
3596191406'), ('Computer Integrated Manufacturing Systems','5.092311859130859'), ('Computer
Technology and Development','5.1069639921188354')]
```

Figure 12. *Recommendation result 01.*

It can be found that according to the input "complex network" and "social computing", the system recommends the top 12 journals with the highest similarity ranking, and gives the similarity values respectively.

(2) Recommendation for the similarity between multi-keywords and all keywords in journals. Firstly, the input keyword is processed by L-BERT model to obtain the vector $\overrightarrow{in}$. When the input is multiple keywords, the input vector is represented by its average vector. Secondly, the Euclidean distance between $\overrightarrow{in}$ and all keyword vectors $\overrightarrow{i_1}, \overrightarrow{i_2}, \ldots, \overrightarrow{i_{5K}}$ in journal are calculated in turn. Finally, the average value is used as the input to calculate the similarity value with the current journal, and the recommended results are arranged in ascending order of the similarity value.

Taking the "Neural Network" and "Machine Learning" as examples, the journal recommendation results are shown in Figure 13:

```
Loading model cost 1.056 seconds .
Prefix dict has been built successfully .
Please enter key words and separated by spaces:  Neural network Machine learning
[('Intelligent Computer and Applications','7.360055685043335'), ('Electronic Technology and Software Engineering','7.369159698486328'), ('Software Guide','7.385413885116577'), ('Modern Computer','7.460778713226318'), ('Computer and Modernization', '7.526304 006576538'),('Computer Simulation','7.557947635650635'), ('Computer Technology and Development','7.62119102478027З'), ('China Computer Communication', '7.642543792724609'),('Electronic Design Engineering','7.643065690994263'), ('Journal of Computer Research and Development','7.646728038787842'), ('Journal of Software', '7.670792818 069458'), ('Information Technology',7.674675464630127')]
```

Figure 13. *Recommendation result 02.*

It can be found that according to the input "neural network" and "machine learning", the system recommends the top 12 journals with the highest similarity ranking, and gives the similarity values respectively.

Multi-dimensional information recommendation. In the recommendation experiment, Multi-dimensional information refers to the title, keyword and abstract information of the paper. Firstly, the input content is processed by word segmentation. Secondly, the words are trained into corresponding word vectors by BERT model, and the average value of multiple word vectors is taken as the representative vector of input items. Finally, the Euclidean distance between the representative vector $\sum_{j=1}^{5K} i_d / 5K$ and all the vectors contained in the journals processed by L-BERT model is evaluated one by one, and the calculation results are sorted in ascending order, which is used as the basis for scholars to recommend journals.

Taking a paper entitled "Research and Application of Keyword Promotion and Advertising in Online Trading Platform" published on Microcomputer Applications as an example, the recommended results based on L-BERT model are shown in the following figure:

```
Please enter the title Research and Application of Keyword Promotion and Advertising in Online Trading Platform
Please enter a summary :It is important to choose the right word in the e-commerce marketplace otherwise it is easy to get lost in the huge platform. The right keywords play an important role in advertising your products on the platform. Sentences for each word provide the location in some search engine page. Therefore, for visible light and position in the web page it is necessary to know the public keywords that competitors sell. This paper proposes a keyword selection algorithm specifically for keyword bidding in the field of e-commerce. This kind of research has strong practical application value. When e-commerce advertisers carry out advertisement promotion, due to the complexity of keyword bidding mechanism, incomplete information characteristics and lack of deep understanding of keywords by advertisers themselves, keywords The bidding strategy of the auction advertisement is wrong. This paper redefines the keyword selection algorithm to adapt to the text characteristics in the field of e-commerce. Obtain recommended keywords through product title text. The whole process first establishes a hierarchical industry-related vocabulary, and then completes the keyword recommendation according to the vocabulary. The recommendation is divided into two steps: seed keyword extraction; recommended keyword expansion.
Please enter key words and separated by spaces :  Recommended keyword expansion search bidding  advertising
[('Microcomputer Applications','9.462924686982669'), ('Application Research of Computers','9.50758628292369'), ('Journal of Chinese Computer Systems','9.546669490834608'), ('Computer Engineering and Applications','9.553 997942470613'), ('Journal of Computer-Aided Design & Computer Graphi','9.589315359330222'), ('Computer Engineering and Design','9.60046565723929 2'), ('Computer Knowledge and Technology','9.607397482356923'), ('Computer Applications and Software','9.624819607658182' ), ('Acta Electronica Sinica','9.651544657620517'), ('Computer Engineering','9.656853808438715')]
```

Figure 14. *Recommendation result 03.*

According to the input information, the system recommends journals based on the similarity between paper and journal, and the similarity values are given respectively. It is found that the first recommended journal is Microcomputer Applications.

4.3. Comparative Experimental

In the experiment, Word2Vec and BERT are respectively used as the training model for vector conversion of experimental data, and at the same time, the "topic-keyword" matrix after LDA training is introduced. All topics contained in the journal and their keywords are trained with the above two language models to obtain the corresponding vectors, and the Euclidean distance is calculated with the vectors obtained after processing the input content, so as to recommend the journals that can be submitted. And the recommendation efficiency is measured by comparing the results of evaluation index values in different situations. Here, the model trained by LDA and Word2Vec is called LDA-Word2Vec model (L-Word2Vec model for short).

Taking "complex network" and "social computing" as the input keywords, the recommendation process and result after the L-Word2Vec model training is shown in Figure 15:

```
Loading model cost 1.056 seconds .
Prefix dict has been built successfully .
Please enter key words and separated by spaces: complex network social computing
[('Software Guide','7.152811527252197'), ('Electronic Technology and Software Engineering', '7.20848822593609'),('Modern Computer','7.256669282913208'), ('Computer and Modernization','7.26054310798645'),('China Computer Communication','7.295558929443359'), ('Intelligent Computer and Applications','7.301868200302124'),('Journal of Computer Research and Development','7.311450958251953'), ('Computer Technology and Development','7.335751056671143'),('Computer Science and Engineering','7..364773988723755'), ('Journal of Software','7.3710105419158936'),('Computer Simulation','7.38300895690918'), ('Electronic Design Engineering','7.401414155960083')]
```

Figure 15. *Recommendation result 04.*

According to the input keywords of "complex network" and "social computing", the system recommends the top 12 journals with high similarity based on L-Word2Vec model. It can be found that compared with the recommendation results trained by L-BERT model, the Euclidean distance of the latter is closer in the same journal, indicating that the effect of L-BERT model is better than L-Word2Vec model.

Taking "Neural network" and "Machine learning" as examples, the journal recommendation results are shown in Figure 16:

```
Loading model cost 1.100 seconds .
Prefix dict has been built successfully .
Please enter key words and separated by spaces: Neural network Machine learning
[('Electronic Design Engineering','13.948364734649658'), ('Software Guide','14.100433826446533'), ('Intelligent Computer and Applications','14.21357774734497'), ('Journal of System Simulation','14.334648609161377'), ('Computer and Modernization','14.370930671691895'), ('Computer Integrated Manufacturing Systems','14.634689807891846'), ('Journal of Zhejiang University-Science A','14.646857738494873'), ('Computer Simulation','14.726103782653809'), ('Computer Systems Applications','14.748590469360352'), ('Computer Technology and Development','14.861507415771484'), ('Computer Science and Engineering','14.882959842681885'), ('Acta Electronica Sinica','14.900765419006348')]
```

Figure 16. *Recommendation result 05.*

Figure 16 shows that when comparing keywords with all keywords in the journal and then recommending according to the similarity, the same content is input, and the recommendation results are quite different after training by L-BERT model and L-Word2Vec model. For journal recommendation of the same content, the Euclidean distance of the former is less than that of the latter. It can be seen that the training effect of L-BERT model is better than the L-Word2Vec model.

Taking a paper entitled "Research and Application of Keyword Promotion and Advertising in Online Trading Platform" published on Microcomputer Applications as an example, the recommended results based on L-Word2Vec model are shown in the following figure:

```
Please enter the title : Research and Application of Keyword Promotion and Advertising in Online Trading Platform
Please enter a summary : It is important to choose the right word in the e-commerce marketplace otherwise it is easy to get lost in the huge platform. The right keywords play an important role in advertising your products on the platform. Sentences for each word provide the location in some search engine page. Therefore, for visible light and position in the web page it is necessary to know the public keywords that competitors sell. This paper proposes a keyword selection algorithm specifically for keyword bidding in the field of e-commerce. This kind of research has strong practical application value. When e-commerce advertisers carry out advertisement promotion, due to the complexity of keyword bidding mechanism, incomplete information characteristics and lack of deep understanding of keywords by advertisers themselves, keywords The bidding strategy of the auction advertisement is wrong. This paper redefines the keyword selection algorithm to adapt to the text characteristics in the field of e-commerce. Obtain recommended keywords through product title text. The whole process first establishes a hierarchical industry-related vocabulary, and then completes the keyword recommendation according to the vocabulary. The recommendation is divided into two steps: seed keyword extraction; recommended keyword expansion.
Please enter key words and separated by spaces: Recommended keyword expansion search bidding advertising
[('Software Guide','10.78551881943109'), ('Electronic Design Engineering','10.86547608645457'), ('Journal of System Simulation','11.063535388910546'), ('Computer Integrated Manufacturing Systems','11.323998716642272'), ('Computer and Modernization','11.3428968978378'), ('Journal of Chinese Information Processing','11.615962109475765'), ('Intelligent Computer and Applications','11.764961526078999'), ('Journal of Zhejiang University-Science A','11.794251136060032'), ('Computer Systems Applications','11.96816766936824'), ('Journal of Computer Research and Development','11.974530795835099')]
```

Figure 17. *Recommendation result 06.*

Figure 17 shows that the system recommends journals according to the input multi-dimensional information. In the figure, the first journal recommended is the Software Guide. Therefore, compared with the recommendation results after L-BERT model training, the multi-dimensional information recommendation results after L-Word2Vec training are poor in similarity.

Please enter the title : Wind power prediction based on artificial neural network

Please enter a summary : The output power prediction of wind farms is of great significance to the operation of number of wind power. Methods for wind speed and wind farm output power prediction are classified. According to the influencing factors of wind farm output power, a neural network model for wind power prediction is established. The effects of measured power data and atmospheric data at different altitudes on the prediction results are analyzed. An error band prediction model based on neural network is established, and the error band prediction is realized. The research results show that the structure of the neural network and the input samples have a certain influence on the prediction results; the measured power data can be used as input to improve the prediction accuracy of the lead time of 30min, while the prediction accuracy of the lead time of 1h will decrease; The prediction accuracy of using all the data as the input of the neural network is higher than that of using only the hub height data; the designed neural network can predict the error band.

Please enter key words and separated by spaces : wind farm power forecast artificial neural network

('Application Research of Computers','9.487124593634354'), ('Microcomputer Applications','9.495099381396646'), ('Computer Engineering and Applications','9.510778590252524'), ('Journal of Chinese Computer Systems','9.547239178105405'), ('Computer Applications and Software','9.573623597621918'), ('Computer Engineering and Design','9.581881422745553'), ('Journal of Computer-Aided Design & Computer Graphi','9.583435824042873'), ('Computer Knowledge and Technology','9.608323517598604'), ('Computer Engineering','9.611937516614011'), ('Modern Electronics Technique','9.614315114523235')]

Figure 18. *Recommendation result 07.*

Taking the paper entitled "Wind power prediction based on artificial neural network" which is not in the database as an example, the recommendation results after multi-dimensional information processing by L-BERT model are shown in Figure 18. It is found that the best journal recommendation result for this paper is Application Research of Computers, and the Euclidean distance is 9.4871. At the same time, the study found that there are a large number of paper data records covering the topics of "artificial neural network" and "wind power prediction" in the computer application research journal, indicating that the recommended results are in line with the expected results.

4.4. Experimental Evaluation Criteria

There is a problem of recommendation efficiency in the recommendation system. The Precision [18], Recall [19] and F1-value [20] commonly used to evaluate the recommendation efficiency. In this paper, the Perplexity is used to select the number of topics in the LDA model in the journal recommendation model, the Precision, Recall and F1-value are used as the evaluation indicators recommended by journals to measure the efficiency of the L-BERT model.

(1) Precision. Precision refers to the proportion of true positive samples in terms of recommended samples. For the prediction results of journal data, it refers to the ratio of the number of correct journals to the total number of recommended journals in the recommended results. The mathematical formula is as follows:

$$\text{Precision}=\frac{TP}{TP+FP} \tag{10}$$

(2) Recall. Recall refers to the proportion of true positive samples that are correctly predicted among the positive samples. That is, there are two possible situations, one is to predict the original positive sample as a positive class (TP), and the other is to predict the original positive sample as a negative class (FN). In the application of journal recommendation, it refers to the ratio of the number of correct journals in the recommended results to the total correct results. The mathematical formula is as follows:

$$\text{Recall}=\frac{TP}{TP+FN} \tag{11}$$

F1-Score. Implementing the recommendation function often needs to expand the scope of the recommendation set to ensure the improvement of the hit rate. At this time, the Recall indicator increases and the Precision indicator decreases. The two indicators will contradict each other and restrict each other. Therefore, it is necessary to comprehensively consider the two indicators for calculation. F-Score is used, which is the weighted harmonic mean of precision and recall, and its mathematical expression is as follows:

$$F=(\gamma^2+1)\times\frac{P\times R}{\gamma^2(P+R)} \tag{12}$$

The γ is the weight factor. When $\gamma=1$, Recall and Precision are considered equally important, that is, F1-value. The mathematical expression is as follows:

$$F1=2\frac{P\times R}{P+R} \tag{13}$$

4.5. Analysis of Experimental Results

In order to comprehensively analyze the recommendation effect of the experimental data in the L-BERT model and the L-Word2Vec model, In this paper, it is agreed that journals with Euclidean distance less than 10 are recommended as positive samples. For example, when the input keyword is "artificial intelligence", the recommendation method of comparing the similarity with the journal vector is adopted, and the recommendation results are shown in Table 5:

Table 5. *Recommended journals and European distance values.*

Recommended journals and European Distance values			
L-Word2Vec		**L-BERT**	
Journal	**European Distance**	**Journal**	**European Distance**
Journal of Chinese Information Processing	8.9460	Electronic Technology and Software Engineering	7.9308
Modern Computer	8.9562	Computer Education	7.9899
Intelligent Computer and Applications	8.9712	Intelligent Computer and Applications	8.0069
Computer Technology and Development	8.9913	Software Guide	8.0365
Chinese Journal of Image and Graphics	9.0016	Modern Computer	8.1668
Computer Science	9.1243	Computer and Modernization	8.3125
Computer Systems Applications	9.4086	Computer Simulation	8.5136
...	...	Electronic Design Engineering	9.2648
Computer and Modernization	9.8911	Computer Engineering	9.6237
Computer Simulation	10.0376	...	...
Software Guide	10.1173	Computer Technology and Development	9.8246
Journal of Chinese Computer Systems	10.1765	China Computer Communication	10.1379
...	...	...	...

The Euclidean distance less than 10 is defined as a positive sample, the number of positive samples in the L-Word2Vec model is 13, and the number of positive samples in the L-BERT model is 14. At this time, set the number of top journals recommended for users to 12. It can be found that the Precision, Recall and F1-value of the recommended results are shown in Table 6:

Table 6. *Measurement indicators and values of recommended results.*

Model	Precision	Recall	F1-value
L-Word2Vec	75%	69%	72%
L-BERT	92%	79%	85%

From the results in Table 6, it can be found that when the input keyword content is "artificial intelligence", the Precision, Recall and F1-value of the recommendation results trained by L-BERT model are 17%, 10% and 13% higher than those trained by L-Word2Vec model. It shows that in the journal recommendation based on scientific research data, the training effect of the L-BERT model proposed in this paper is better than the traditional L-Word2Vec model.

5. Conclusion

This paper proposes a journal recommendation method based on the L-BERT model, and uses 36 types of computer-related journals with a total of 21,700 paper records as the experimental dataset to conduct journal recommendation experiments.

By comparing the recommended results after training the experimental data on the L-BERT model and the benchmark model L-Word2Vec, it found that the Precision, Recall and F1-value of the data trained by L-BERT model are increased by 17%, 10% and 13%, respectively. It shows that the recommendation effect of this model is better than the benchmark model on the basis of experimental data.

However, the experimental data in this paper only contains data records of 36 types of journals related to computers in the past ten years, and the dataset used are limited in breadth and quantity. In the following work, subsequent scholars can expand the experimental data set, expand the scope of data collection from a single computer field to multiple professional fields, and integrate larger-scale data for experimental analysis. And they can add other comparative models based on the benchmark model in this paper to explore more experimental possibilities.

Acknowledgements

The paper was supported by the National Natural Science Foundation of China (No. 61763041), the Science and Technology Department of Qinghai Province (No. 2022-ZJ-T02).

References

[1] Wu Z, Wu J, Khabsa M, et al. Towards building a scholarly big data platform: Challenges, lessons and opportunities [C]// IEEE/ACM Joint Conference on Digital Libraries. IEEE, 2014: 117-126.

[2] Rossmann, Bernhard, Canzaniello, et al. The future and social impact of Big Data Analytics in Supply Chain Management: Results from a Delphi study [J]. Technological forecasting and social change, 2018, 130 (May): 135-149.

[3] Ahmed M M, Kader M A, Zamli K Z. Comparative Study to Measure the Quality of Big Scholarly Data and Its Hypothetical Mapping Towards Granular Computing [J]. Journal of Computational and Theoretical Nanoscience, 2018, 24 (10): 7810-7814.

[4] Wang D, Liang Y, Xu D, et al. A content-based recommender system for computer science publications [J]. Knowledge-Based Systems, 2018, 157: 1-9.

[5] Wang L C. Research on Paper Recommendation System Based on Deep Neural Network [D]. Jilin University, 2019.

[6] Yang Z H, and Brian D. Davison. "Venue Recommendation: Submitting Your Paper with Style." 1 (2012): 681-686.

[7] Dai J. Research and Implementation of Cooperative Recommendation Model Based on LDA in Scientific Research Achievement Management System [D]. Liaoning University, 2017.

[8] Zhang Y, Dong L L, Li S H. Extraction of Elements of Protest Based on BERT Model and TextTeaser improved algorithm [J]. Journal of Physics: Conference Series, 2021, 1955 (1).

[9] DEVLIN J, CHANG M W, LEE K, et al. BERT: Bidirectional encoder representations from transformers for language understanding [J]. Computation and Language, 2018, 23 (2): 3-19.

[10] Zhang Y, Hao J B. Malicious code detection method based on attention mechanism and residual network: 1-10, 2022.

[11] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space [J]. CoRR, 2013, abs/1301.3781.

[12] Fan W, Wang H M, Xing Y. Auto-encoder based multi-view attributed network representation learning model [J]. Journal of Computer Applications, 2021, 41 (04): 1064-1070.

[13] Liu S Z, Zhu Y D, Chen R Z, Luo H Y, Zhao F, Sun Y, Wang B H. Traffic mode recognition algorithm based on residual temporal attention neural network [J]. Journal of Computer Applications, 2021, 41 (06): 1557-1565.

[14] Liu H L, Ma J, Zhang G X. Review of studies on deep learning-based content recommendation algorithms [J]. Computer Engineering, 2021, 47 (7): 1-12.

[15] Rajesh T M, Kulkarni Praveen. A Multi-Model Framework for Grading of Human Emotion Using CNN and Computer Vision [J]. International Journal of Computer Vision and Image Processing (IJCVIP), 2021, 12 (1).

[16] Zhou C, Li X L, Li Q L, Zhu D D, Chen S L, Jiang L R. Sparse Non-negative Matrix Factorization Algorithm Base on Cosine Similarly [J]. Computer Science, 2020, 47 (10): 108- 113.

[17] Qiu J P, Shen C. Analysis of Hot Topics in Domestic Big Data Research Based on LDA Model [J]. Journal of Modern Information, 2021, 41 (09): 22-31.

[18] Bao X, Chen H M, Xiao Q. Time-incorporated point-of-interest collaborative recommendation algorithm [J]. Journal of Computer Applications, 2021, 41 (08): 2406-2411.

[19] Yang X L. Survey for Performance Measure Index of Classification Learning Algorithm [J]. Computer Science, 2021, 48 (08): 209-219.

[20] Dong J W, Shi Y Q, Xie B R. Adaptive Recommendation System for Personalized Learning Resource Based on Artificial Intelligence [J]. China Computer & Communication, 2021, 33 (15): 166-168.