Hindawi
Journal of Sensors
Volume 2022, Article ID 9953787, 8 pages
https://doi.org/10.1155/2022/9953787

Research Article

Digital Library Information Integration System Based on Big Data and Deep Learning

Xiao Lin, Ying Zhang, and Jiangong Wang

The Minjiang University Library, Fuzhou, Fujian 350000, China

Correspondence should be addressed to Xiao Lin; 1401040237@xs.hnit.edu.cn

Received 19 May 2022; Revised 4 June 2022; Accepted 15 June 2022; Published 30 June 2022

Academic Editor: C. Venkatesan

In order to solve the defects of traditional text classification in digital library, the author proposes a method based on deep learning in the field of big data and artificial intelligence, which is applied to the digital library information integration system. On the basis of systematically sorting out the traditional text classification of digital library of this method, this paper proposes a digital library text classification model based on deep learning and uses the word vector method to represent text features, the convolutional neural network in the deep learning model is used to extract the essential features of text information, and experimental verification is carried out. Experimental results show that deep learning-based text classification model can effectively improve the accuracy (average 94.8%) and recall (average 94.5%) of text classification in digital libraries; compared with the traditional text classification method, the text classification method based on deep learning improves the average F1 value by about 11.6%. *Conclusion*. This method can not only improve the intelligence of the internal business of the digital library, but also improve the efficiency and quality of the information service of the digital library.

1. Introduction

With the rapid and organic integration of computer, network technology, communication technology, and other technologies, digital information has been widely used in the human environment. Digital and other forms of information have a life cycle and play a role in that life cycle [1]. Its life cycle includes production, storage, analysis, distribution, change, innovation, and reproduction. The digital library has rich digital information resources, such as e-books, images, audio and video materials, CD-ROMs, and other media. It has a technology platform that can provide advanced simple and fast information services such as intelligent messaging, data transfer, and mobile services.

The emergence of the digital library is inseparable from the popularization of the Internet. The digital library is the crystallization of human wisdom. It has changed the way people obtain information. Make it easier and more colorful for people to acquire knowledge. The digital library is not only a new development in science, but also a new branch of public electronics [2]. Computer technology, network storage technology, communication technology, and many other technologies have developed rapidly in the past ten years. Their development and expansion have expanded a lot of digital information, and now people are moving around digital information. In fact, due to the widespread use of digital information, the way people receive information has changed dramatically, and people no longer want to receive information through text, but transmit electronic data through the Internet [3]. Due to the huge amount of Internet data, how to obtain more, better, more accurate, timely, and useful information from it has become a problem that people care about.

2. Literature Review

Wang, J. et al. discussed the application prospects of expert systems in libraries [4]. Shi, M. et al. describe how a knowledge-based librarian system, UMLS, can be used to search the MEDLINE bibliographic database [5]. ∗Wu, Y. et al.'s research on AI technology in digital libraries, specifically discussed how to use Daubechy wavelet transform for image indexing; An extension of the Stone Li algorithm for handling occlusions in images; Fault-tolerant structure

extraction strategies for semistructured text documents, etc. [6]. According to Kim, Y. et al., the structure of each department responsible for data retrieval in the digital library is described. Ontology is used to present problems, avoid ambiguous information as much as possible, and also use ontology to identify relevant information [7]. The formation of these agents is based on the Gaia method. According to Wang, D. et al., some applications of fuzzy light theory in retrieving data are reported, as well as final research in the field [8]. Lu, L. et al. show how to better answer example questions posed using static electronics such as blinking and gazing [9]. Yus, Y. et al. explore the development and use of international conferences while focusing on the availability of Canadian libraries [10].

At present, for the in-depth research of library data service users, data recovery, and intelligent question answering robot, there are also some researches on intelligent research of digital library business information distribution. At the same time, some scholars from many countries discussed the application of in-depth research in the field of psychology, but most of them focus on using deep learning for entity extraction, information extraction, cross-language retrieval, and sentiment analysis; however, there are few related researches on the application of deep learning in library text classification, automatic summarization, and topic extraction.

This paper summarizes the traditional text classification methods from two aspects: text feature representation and text feature selection. He researches and writes about text distribution patterns. Traditional text categories do not have high-dimensional sparse data or semantics, so text classification is recommended. The library model first converts the data files in the digital library, and then uses the word vector method to represent the text as two-dimensional network data, which can solve the traditional text representation method. Latitude problem, sparse data, no semantics, with word vector as the concept, in-depth study of recurrent neural network model. Through the special design of the convolutional neural network, the basic feature is that it can delete the text and finally complete the text to achieve the purpose of improving the text quality of the digital library.

3. Research Methods

3.1. Overview of Traditional Text Classification. Text classification in digital library refers to the classification of text according to the subject, content, or attribute of text information under the premise of a given classification standard and the process of classifying massive text information resources into single or multiple categories. As an important basis for information management and organization in digital libraries, text classification can help users find and locate the required information quickly and accurately, and it is a very effective method to manage textual information resources with huge amount of data [11].

The text classification process in a traditional digital library includes four parts, namely, text preprocessing, feature representation, feature selection, and classifier, of which the most important are text feature representation and text feature selection, which play a decisive role in the results of text classification [12]. The main process of text classification is shown in Figure 1.

3.1.1. Text Feature Representation. Special representation is the basis of text distribution in digital library and an important function of information organization and management in digital library, and a good text feature representation plays a decisive role in the performance of text classification tasks in digital libraries. Currently, scientists have proposed a variety of esthetic models: Boolean models, probabilistic models, vector space models, and various hybrid models [13]. Among them, vector space modeling has been widely used in recent years and is one of the most effective methods to describe text features.

The vector space model is a text feature representation method based on statistical theory, this representation method is simple and direct, compared with other models, the text feature representation method of the vector space model is also more standardized, and the use effect is also ideal. Therefore, the vector space model has always been the focus of scholars. Although the vector space model can reduce the complexity of text processing and improve processing efficiency to a certain extent, however, it leads to the high dimensionality of text feature vectors and the consequent data sparse phenomenon. In response to this problem, the first thing is to find a method that can reduce the dimension of text features without affecting the effect of text classification, and this is the dimensionality reduction process that people need to perform before classifying.

3.1.2. Text Feature Selection. Reducing the dimension of text functions is a key factor affecting the accuracy and efficiency of text classification in digital libraries. Currently, there are two main ways to reduce script size: select specific and remove features. Feature selection is the basis for ignoring the first feature. From the initial text settings, select some features that best represent the text content and then reduce the size of the text features. Improve the accuracy and efficiency of text distribution in digital libraries. The most common options include data frequency, chi-square statistics, and data gain [14].

Compared to optional features, feature extraction takes a more efficient way, shifting the importance of the text source from the art to creating a new feature source of shorter length, longer duration, and more freedom to improve the text and dimensionality reduction. However, since the subtractive features are very close to identifying semantic scripts and the technology involved in this is immature, the impact size is not good. Compared with feature extraction methods, script features for selection feature selection are a subset of the original source features and have the advantages of semantic context, easy understanding, simple structure, and ease of use. Therefore, it has attracted the attention of many scientists and has become an important means of reducing the size.

3.2. Text Classification Based on Deep Learning. After analyzing the characteristics of traditional text classification

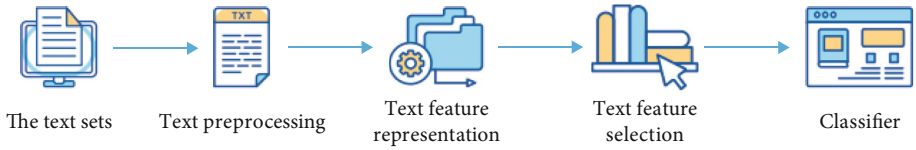

Figure 1: Text classification process.

methods, it will be found that when dealing with the problem of digital library text classification in the big data environment, there are many difficult problems to solve. In the process of feature extraction, manual participation is required, and it will affect the accuracy of the final extracted text features. The use of vector space models in text representation ignores the semantic and semantic information in the text, which affects text features; in the case of high-dimensional and sparse data, although attribute selection methods can be used to reduce dimensionality, it increases feature and data loss risks, making text processing more complex overall [15]. There are few features that do not affect the distribution of the text distribution process, Joachims said, and a good text distribution standard should take advantage of all features. Therefore, in order to complete the text work of the digital library in the era of big data, it is necessary to make the text more efficient, high quality, and higher in text design.

The authors' in-depth courses based on classification models for digital libraries include pre-written scripts, vector-based word representations, pamphlets, and classification using convolutional neural networks [16]. Text preprocessing uses word vectors to represent text, and the convolutional neural network in the deep learning model is used to extract text features and complete the final classification. The main process of deep learning-based digital library text classification is shown in Figure 2.

3.2.1. Text Preprocessing. Therefore, before the text classification, the original text information needs to be preprocessed; in this way, the subsequent text classification tasks can be carried out. The quality of text preprocessing has a great influence on the accuracy of text classification results. Text preprocessing mainly includes text segmentation and stop word removal.

(1) *Participants*: Characters, words, and phrases in Chinese appear in sequence without specific segmentation. A word is the basic unit of meaning. When sorting text, a word is needed to describe the characteristics of the text. Therefore, Chinese word segmentation is a relatively basic and important link in text classification. At present, the word segmentation technology is relatively advanced, and word segmentation tools such as jieba word segmentation, Baoding analyzer, and ICTCLAS of the Chinese Academy of Sciences are widely used [17]

(2) *Go to the station*: Delete words that are meaningless but frequently appear in the text. These words are not helpful for segmenting text, as they do not represent text, but increase the size of the text vector. This will affect the final effect of text classification; therefore, it needs to be removed so that the remaining text information can better express text features

3.2.2. Text Feature Representation: Word Vectors. When using deep learning to perform text classification tasks in digital libraries, it is first necessary to use text representation methods and convert semistructured or unstructured text into vector representations that computers can understand and process. In view of the problem of high latitude of vectors, data is scarce, there is no semantics in the traditional text distribution process, and the benefits of digital libraries are not good; the author uses the word vector method to represent the text content, which can effectively solve these problems.

The traditional text representation model represents the text features, which will make the text feature space very high, and the deep learning model cannot exert its powerful feature extraction ability when classifying it. With the proposal of word vector method, the text feature representation based on word vector provides a prerequisite for deep learning to be used for text classification tasks in digital libraries.

A word vector is a combination of each word in the text by plotting each word in the text as a constant of small dimension in real space. Instead of expressing multiple real spaces by separating vector space models, high-dimensional sparse vectors can be transformed into low-dimensional, dense real vectors.

Compared with the usual representation card, after the word vector is mapped in the new low-dimensional text feature position, the relationship between the word vector relative to words with different letters represents the relationship between texts, and it can provide richer textual semantic information. At the same time, the word vector method can overcome the problems of high vector dimension and data sparseness in grammar models. This is a big advantage that text representation does not have, so using word vector method to represent text can improve the quality and accuracy of distributed literature in digital libraries [18].

Word2vec is a tool for training and designing word vectors launched by Google in 2013 based on in-depth research. It can train a large amount of corpus conveniently, and through training, it can effectively map the feature words in the text into dense vector at low latitude.

3.2.3. Text Feature Extraction. In the context of the previous section, the use of word vectors to represent text solves the problem of text representation. This demonstrates the use of convolutional neural networks in deep learning standards to solve the problem of automatic subtraction of text distributions in digital libraries. As one of the most deeply researched

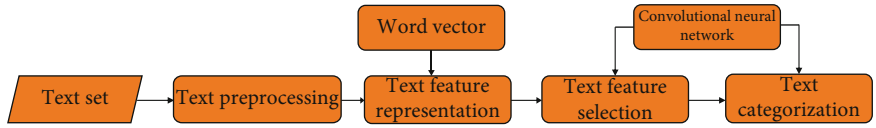

Figure 2: Text classification process based on deep learning.

models, neural communication has been used by many scientists to solve problems in natural language processing, and writers have attempted to use these connected neural languages to enable the distribution of text in digital libraries.

The basic structure of the convolutional neural network is shown in Figure 3. The first half consists of the input layer, and the middle part consists of the multirecurrent layer and the consolidation layer; i.e., an additional recurrent layer is added. The second half consists of a generic connection layer and an output layer, where one layer is connected first and then to the recurrent layer [19].

The training of convolutional neural networks can be thought of as a healing problem, and the functional objective should be designed to be maximized or reduced according to the healing objective. In in-depth research, job cost is often used as a job objective by proper terminology, and job cost is the error in measuring the distribution of objects. Calculate the result and the actual value; the operating cost is usually used in the division function and is the average value of the cross-entropy loss function of each model. Cross-entropy is a measure of the similarity of two probability distributions, and the target distribution is expressed as $p(x)$, the distribution obtained by prediction estimation is denoted as $q(x)$, and the cross-entropy between them is defined as the following formula (1) [20]:

$$H(p, q) = -\sum_{x} p(x) \log q(x). \tag{1}$$

If the label value is represented by a one-hot vector, that is, the label value for k-category classification is represented as a target vector of length k: $[p^1, \cdots, p^j, \cdots, p^k]$; if the target class is $y_i = c$, then let $p^c = 1$, all other items are 0, and then, the final objective function can be expressed as formula (2). The expression $1\{c = yi\}$ indicates that the condition in the parentheses is satisfied, and then take 1; otherwise, it is 0.

$$L = -\frac{1}{m}\sum_{i=1}^{m}\sum_{c=1}^{k} 1\{c = yi\} \bullet \log \frac{e^{z_c}}{\sum_{j=1}^{k} e^{z_j}}. \tag{2}$$

The convolution process can eliminate the different local features of the text, and the lower-level convolution process can eliminate the lower-level features of the text such as words, phrases, sentences, lines, and words, and the high-level convolution process can be omitted. High-level features of text, such as sentences, phrases, semantics, etc. The background of the convolutional layer is the pooling layer, which can give semantically similar text features and play the role of secondary feature extraction. The combination of convolution process and layer pooling process is based on the whole ensemble process, and the role of the whole connection process is to identify and assign various local concepts. The function of the full connection layer is to summarize and classify various local text features extracted from the convolutional layer and the pooling layer.Similar to the function of classifiers in traditional text classification, classification information is finally obtained through the output layer.

The special network structure of convolutional neural network makes it have the following characteristics: ① The special structure of the convolutional neural network makes it good at processing grid-type data. Convolutional neural networks can prove their ability to eliminate images and speech, which is why image and speech data are meshed. Data is the difference between images and speech, data is the solution, and the text represented by the model is written as a representation of the negated text using a convolutional neural network. By using the word vector method to represent the text features, the text features can be converted into continuous and dense two-dimensional grid data similar to images and speech, which can be well-processed by the convolutional neural network. ② The combination of convolutional process and layer pooling specially built for convolutional neural networks can exclude locally unreliable features at various levels in the text. Compared with the design structure in the traditional text, the convolutional neural network can learn the characteristics of the data file, which not only saves time and effort, but also avoids the negative impact of the accumulation of errors caused by manual feature extraction. The extracted features are stronger than discrimination ability. ③ When using convolutional neural network to classify text, the feature extraction and classification of text are taken as a whole. In traditional text classification methods, the feature extraction part of text and the classifier are two independent parts. In the convolutional neural network, the feature extraction and classification of text are carried out in the convolutional neural network; as the input of the convolutional neural network, the word vector will be continuously optimized with the training of the network; and therefore, the performance of the joint collaboration of feature extraction and classification as a whole can be maximized, thereby further improving the effect of text classification.

Convolutional neural networks can simplify the process by the aggregation and stacking of layers and layers, which can extract higher-level and more complex solutions, which can render text more intensely, affecting the main features. Convolutional neural network can deepen the number of layers through the accumulation and superposition of convolutional layer and pooling layer, so as to extract higher-level and more abstract text features, so that text features have stronger expression ability and more accurately reflect the essential features

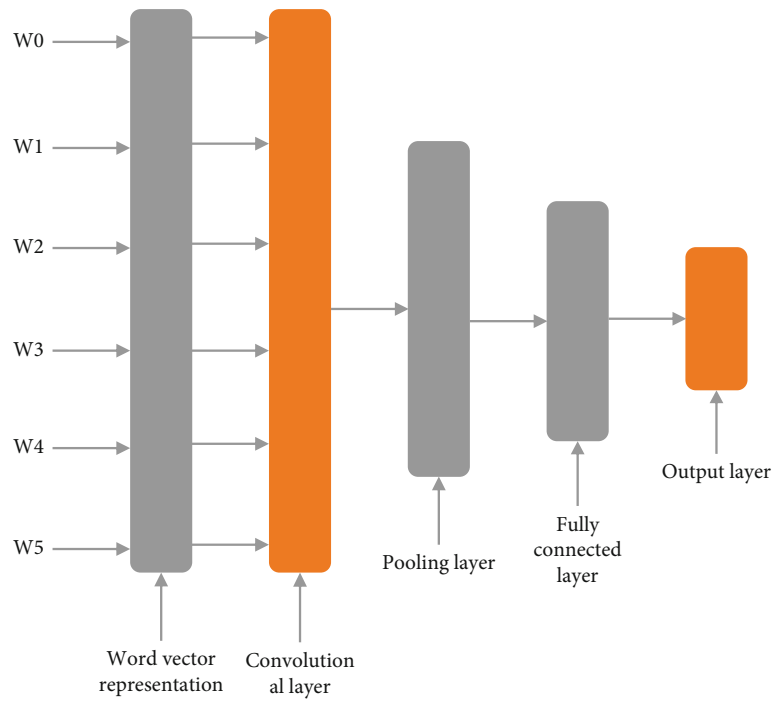

Figure 3: Structure diagram of text classification based on convolutional neural network.

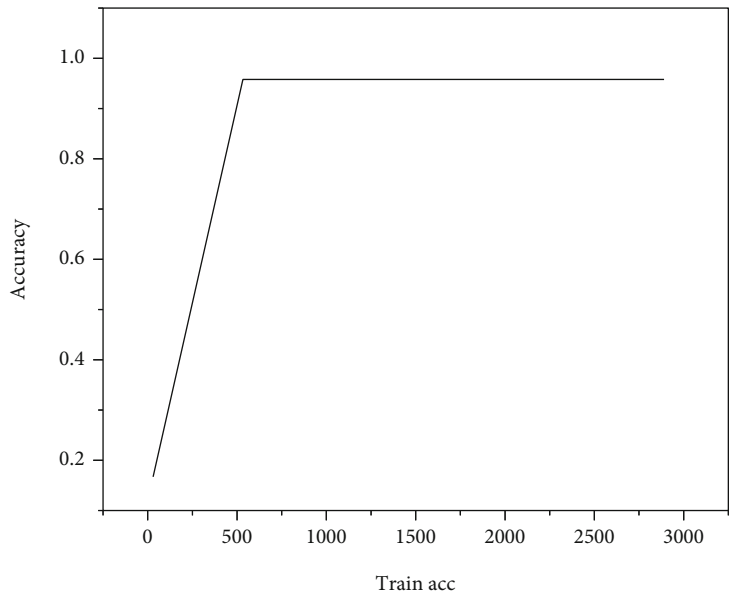

Figure 4: Error rate and accuracy rate of model training set.

of text. While the depth of the layers can be more complex and time-consuming, this can be compensated by increasing hardware performance. Therefore, convolutional neural networks are used to solve the text features of digital libraries, which can improve the performance and accuracy of distribution.

4. Analysis of Results

The authors conducted experiments to determine the distributional advantages of a depth-based digital library. Using convolutional neural network, the experiment of Chinese distribution is carried out through Google's open source deep learning TensorFlow [21]. The most common measures we use in text distribution are to measure the benefits of text distribution: precision, recall, and F1 value.

The experimental data is selected by the author from a set of public records set up by a school's natural language laboratory. The experimental data consists of 10 categories: sports, finance, real estate, real estate, education, technology, fashion, current affairs, sports, and entertainment. There are 6500 pieces of information in the category, and the

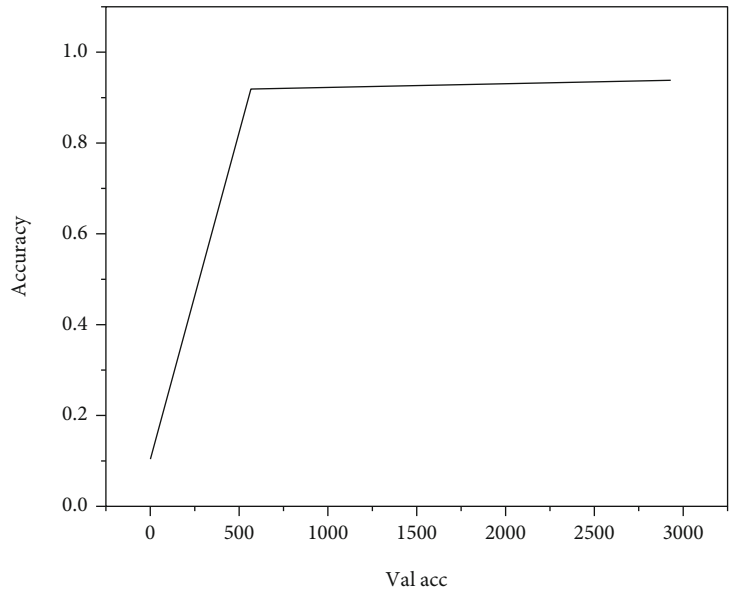

Figure 5: Error rate and accuracy rate of model validation set.

Table 1: Statistics of experimental results of text classification.

Category	Traditional text classification methods			Deep learning-based text classification		
	Accuracy (P)	Recall (R)	F1	Accuracy (P)	Recall (R)	F1
Entertainment	0.78	0.79	0.78	0.94	0.96	0.95
Game	0.83	0.83	0.83	0.96	0.94	0.95
Current affairs	0.86	0.81	0.84	0.96	0.91	0.93
Fashion	0.86	0.87	0.87	0.93	0.96	0.94
Technology	0.86	0.85	0.85	0.95	0.96	0.96
Educate	0.77	0.76	0.76	0.9	0.92	0.91
Furniture	0.87	0.82	0.85	0.95	0.9	0.93
Real estate	0.87	0.89	0.88	0.98	0.96	0.97
Finance	0.82	0.84	0.83	0.94	0.97	0.96
Physical education	0.81	0.83	0.82	0.97	0.97	0.97
Average value	0.833	0.829	0.831	0.948	0.945	0.947

information layers are divided as follows: $5000 * 10$ for training process, $500 * 10$ for usability, and $1000 * 10$ for testing process.

In the preliminary data set, the data experiments were segmented using the ICTCLAS term segmentation system of the Chinese Academy of Sciences. In the process of text representation, the word2vec tool is used to identify the vector representation of text data, and the convolutional neural network model and the convolutional neural network model are studied with the text receiving vector content as the input. And complete the TensorFlow platform test.

Before using the training process and the verification process to train the convolutional neural network, after the training process accuracy rate is stable, it is above 95%, as shown in Figure 4, and the verification process accuracy rate is stable at about 94%, as shown in Figure 5; the text distribution and the performance on training and implementation are very good.

Finally, the distribution results of the convolutional neural network distribution model are identified by a test procedure not included in the training. And vector format is always available for text feature representation, and text selection using TF-IDF is used to compare distributions. The comparative experimental results are shown in Table 1, and the F1 values of the two text distribution methods are compared, as shown in Figure 6. An in-depth study of the average accuracy and average return is shown in Table 1 and Figure 6. According to the text distribution about 94.8% and 94.5%, respectively, the average accuracy and average return of the text distribution process are usually between 83.1% and 82.9%, especially for the deep learning-based text distribution, and the accuracy and the return of the distribution results are very good.

The F1 value of entertainment, games, technology, education, finance, and sports has increased by a large margin, while the F1 value of current affairs, fashion, furniture, and real estate

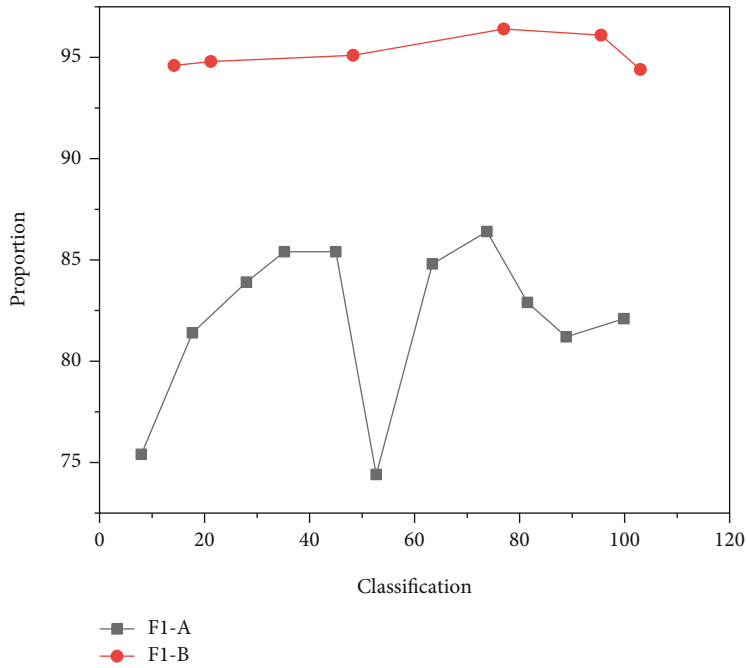

Figure 6: Comparison of F1 values.

has increased relatively little; compared with the traditional text classification method, the text classification method based on deep learning improves the average F1 value by about 11.6%. Therefore, in general, according to the well-studied text distribution used in this paper, it can improve the results of text distribution compared to the text distribution model. This shows that text is represented by word vectors, and then, the neural network connections in the deep learning model are used to extract and segment the text, which can help improve the efficiency of book distribution in digital libraries.

5. Conclusion

On the basis of analyzing the text distribution model of digital library, a text distribution model of digital library based on depth and text model is proposed, while word vectors can carry the syntactic and semantic text of the text, a special hierarchical model connected by a neural network of deep learning models, and features at the syntactic and semantic levels, such as words, phrases, phrases, and sentences. Texts can be taught by themselves. Combining texts with word vectors and convolutional neural networks can work together on the text distribution of digital libraries, which can solve the problems that have always existed in texts. The experimental results show that, compared with the standard text distribution model based on the vector space model, the depth-based digital library text distribution model can meet the author's requirements, and the distribution effect is better.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The second batch of science and technology projects of Fuzhou Science and Technology Bureau: digital library knowledge service platform based on WeChat official account's open source big data (project number: 2021-SG-039).

References

[1] L. Gu, "Integration and optimization of ancient literature information resources based on big data technology," *Mobile Information Systems*, vol. 2021, no. 3, Article ID 6452418, 8 pages, 2021.

[2] Q. Jia, "Research on mobile learning in a teaching information service system based on a big data driven environment," *Education and Information Technologies*, vol. 26, no. 5, pp. 6183–6201, 2021.

[3] X. Lv and M. Li, "Application and research of the intelligent management system based on internet of things technology in the era of big data," *Mobile Information Systems*, vol. 2021, no. 16, Article ID 6515792, 6 pages, 2021.

[4] J. Wang, "Massive information management system of digital library based on deep learning algorithm in the background of big data," *Behaviour and Information Technology*, vol. 5, pp. 1–9, 2020.

[5] M. Shi, "Knowledge graph question and answer system for mechanical intelligent manufacturing based on deep learning," *Mathematical Problems in Engineering*, vol. 2021, no. 2, Article ID 6627114, 8 pages, 2021.

[6] Y. Wu, G. Tian, and W. Liu, "Research on moisture content detection of wood components through wi-fi channel state information and deep extreme learning machine," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 9977–9988, 2020.

[7] Y. Jin, G. Li, and J. Wu, "Research on the evaluation model of rural information demand based on big data," *Wireless Communications and Mobile Computing*, vol. 2020, no. 5, Article ID 8861207, 14 pages, 2020.

[8] D. Wang and H. H. Lee, "Research on big data privacy protection based on the three-dimensional integration of technology, law, and management," *The Journal of Korean Institute of Information Technology*, vol. 19, no. 3, pp. 129–140, 2021.

[9] L. Lu and J. Zhou, "Research on mining of applied mathematics educational resources based on edge computing and data stream classification," *Mobile Information Systems*, vol. 2021, no. 7, Article ID 5542718, 8 pages, 2021.

[10] Y. Yu, "Retracted article: numerical simulation of sea surface temperature based on big data and calculation of economic effect of import trade," *Arabian Journal of Geosciences*, vol. 14, no. 16, pp. 1–13, 2021.

[11] M. Khosravy, K. Nakamura, Y. Hirose, N. Nitta, and N. Babaguchi, "Model inversion attack: analysis under gray-box scenario on deep learning based face recognition system," *KSII Transactions on Internet and Information Systems*, vol. 15, no. 3, pp. 1100–1119, 2021.

[12] C. Li and J. Cui, "Intelligent sports training system based on artificial intelligence and big data," *Mobile Information Systems*, vol. 2021, no. 1, Article ID 9929650, 11 pages, 2021.

[13] A. I. Tikhonov, A. A. Sazonov, and I. Kuzmina-Merlino, "Digital production and artificial intelligence in the aircraft industry," *Russian Engineering Research*, vol. 42, no. 4, pp. 412–415, 2022.

[14] L. Liu and S. B. Tsai, "Intelligent recognition and teaching of english fuzzy texts based on fuzzy computing and big data," *Wireless Communications and Mobile Computing*, vol. 2021, no. 1, Article ID 1170622, 10 pages, 2021.

[15] R. Huang, "Framework for a smart adult education environment2015," *World Transactions on Engineering and Technology Education*, vol. 13, no. 4, pp. 637–641, 2015.

[16] K. Wei, W. Kong, and S. Wang, "Integration of imaging genomics data for the study of Alzheimer's disease using joint-connectivity-based sparse nonnegative matrix factorization," *Journal of Molecular Neuroscience*, vol. 72, no. 2, pp. 255–272, 2022.

[17] P. Ajay, B. Nagaraj, and J. Jaya, "Bi-level energy optimization model in smart integrated engineering systems using WSN," *Energy Reports*, vol. 8, no. 2490-2495, pp. 2490–2495, 2022.

[18] S. Liu, Y. Yang, and Y. Wang, "Integration of museum user behavior information based on wireless network," *Mobile Information Systems*, vol. 2021, no. 5, Article ID 6847144, 8 pages, 2021.

[19] S. Kannan, G. Dhiman, Y. Natarajan, A. Sharma, and M. Gheisari, "Ubiquitous vehicular ad-hoc network computing using deep neural network with iot-based bat agents for traffic management," *Electronics*, vol. 10, no. 7, p. 785, 2021.

[20] S. Bratulic, F. Gatto, and J. Nielsen, "The translational status of cancer liquid biopsies," *Regenerative Engineering and Translational Medicine*, vol. 7, no. 3, pp. 312–352, 2021.

[21] L. Xin, M. Chengyu, and Y. Chongyang, "Power station flue gas desulfurization system based on automatic online monitoring platform," *Journal of Digital Information Management*, vol. 13, no. 6, pp. 480–488, 2015.