International Journal of Applied Engineering & Technology

A New Method of Text Classification Based on Recurrent Neural Network

Deageon Kim
Architectural Engineering, Dongseo University, 47 Jurye-ro, Sasang-gu, Busan, Republic of Korea
gun43@hanmail.net

Date of Submission: 20th November 2022 Revised: 29th December 2022 Accepted: 10th January 2023

How to Cite: **Kim, D., (2023). A New Method of Text Classification Based on Recurrent Neural Network. International Journal of Applied Engineering & Technology 5(1),pp 13-23.**

Abstract: With the development of modern information science and technology, the number of Internet users continues to increase substantially, and the processing of massive data is now a hot spot in data research. Artificial Neural Network (ANN) plays a crucial role in the screening and processing of big data. Artificial neural network has successfully solved many practical problems that have puzzled people for many years in the fields of computer vision, machine translation, automatic driving, etc. Therefore, artificial neural network has been increasingly applied to text classification in Natural Language Processing, NLP), which is a hot and difficult point in NLP at present. Using artificial neural network can not only process massive data quickly and efficiently, but also improve the accuracy of data processing to a certain extent. However, there are many differences between English and Chinese in character level and word level. Compared with English, the number of Chinese in character level and word level is larger than that of English. At present, the Chinese text classification technology still has some problems in processing speed, accuracy and word segmentation. In this paper, some contents of THUC News data set compiled from Sina news data are extracted for text classification. Firstly, a word embedding matrix based on character level is proposed, and the vector dimension of each word is only 13 dimensions. Through experimental comparison, it is found that the word vector based on character level proposed in this paper has better classification effect than the word vector trained by word2vec. A tower-shaped three-layer bidirectional network structure based on LSTM(Long Short-Term Memory) is also designed, which is connected to the full connection layer composed of three layers of DNN(Deep Neural Networks) networks. Through experimental comparison, it is found that the network model designed in this paper achieves better effect than Text CNN network. In the aspect of convolutional neural network, this paper puts forward a weight compensation scheme to solve the problem that convolutional kernel ignores edge information and extends it to higher dimensions, and puts forward an optimized pooling structure to solve the problem of erroneous information extraction caused by traditional maximum pooling and average pooling. Through experiments, the optimized pooling designed in this paper is superior to the maximum pooling in training convergence speed and classification effect. In addition, a convolution neural network with five parallel connections based on one-dimensional convolution is designed. Through experimental verification and analysis, the parallel convolution neural network designed in this paper achieves better results than the Text CNN network structure. Finally, a CRNN (Convective Recurrent Neural Networks) network which combines CNN (Convective Neural Networks) and RNN(Recurrent Neural Network) is designed. Through experimental verification and analysis, the text classification effect is also better than that of Text CNN network structure.

Keywords: Neural Network, Natural Language Processing, Text Classification, THUC News, RNN, CNN

1. INTRODUCE

1.1 RESEARCH BACKGROUND AND SIGNIFICANCE

Nowadays, with the vigorous development of information science and technology, the number of Internet users continues to grow by a large margin, and the era of big data has quietly arrived. Big data has made great achievements in many disciplines, and for individuals, each of us will face ten times and hundreds times the amount of data in the past every day. However, there are many problems in processing speed, processing accuracy, resource cost and so on when using traditional methods to process massive data. Therefore, the appearance and development of Artificial Neural Network (ANN) provide a faster and more accurate way to deal with massive data to a certain extent [1]. Artificial neural network is a mathematical mapping model established by observing and imitating the animal neural network in nature, combining the brain science features at the biological level, extracting the corresponding features of massive data, and then building a corresponding network structure similar to biological neural network, and performing corresponding nonlinear operations on the features of data [2]. The internal operation process of the neural network is extremely complicated, and there is a certain black box characteristic. It mainly processes the relevant information by constantly adjusting the internal neurons, and the neural network model has the ability of autonomous learning [3]. Text classification refers to the process of dividing a large number of related texts into a number of related but different categories according to certain standards under a certain classification system. Text classification is a research hotspot in natural language processing [4]. Before the 1990s, the commonly used text classification methods were always based on manual classification, that is, the relevant professional researchers manually classified the texts. However, manual classification is not only inefficient and costly, but also has non-objective problems caused by physical strength and other factors in classification accuracy. After the 1990s, with the introduction of many classification methods and machine learning, a new research direction of text classification technology has emerged. The rapid development of text classification technology has attracted many researchers, and the text classification technology has made great achievements in English text classification [5]. On this basis, the domestic research on Chinese text classification has also been developed, and some achievements have been made in many fields, such as book management, web page classification, information statistics, quick search and so on. Among the neural networks, convolutional neural networks (CNN) can capture all parts of local information in the text data, and recurrent neural networks (RNN) pay more attention to the context and semantic relations in the text, so these two kinds of neural networks are often applied to the field of text classification and have made corresponding development [6-10]. However, although the application of neural network in text classification has made a series of achievements, there are still some problems to be solved [11]:

（1）Word segmentation text classification includes character-based and word-based classification, while the number of Chinese is much larger than that of English in character-based and word-based, which leads to the difficulty of Chinese text classification to a certain extent. Excellent word segmentation can significantly improve the training speed and classification accuracy of text classification.

（2）Feature extraction and feature selection Both feature extraction and feature selection are aimed at reducing the dimension of feature vectors. Different feature extraction and feature selection methods have great influence on text classification, and are a very important step in text classification. In the current text classification model, there is still great room for optimization in feature extraction and feature selection.

（3）Convolutional neural network pool layer The existence of pool layer in CNN will cause us to lose a lot of information in the process of processing information, and at the same time, it can't handle the relationship between local and whole well. For example, when we use convolutional neural network to recognize the text, we recognize the whole text by recognizing the features of each part of the text, but ignore the correlation between each part and the whole text [12].

（4）Deep neural network is difficult to train. Theoretically, compared with shallow neural network, deep neural network is more excellent in performance. However, the training of deep neural network is really very difficult. The training of neural network needs to use vectorization to replace the for loop as much as possible in order to give full play to the computing power of CPU or GPU. Therefore, training neural network requires higher hardware cost. Moreover, if the number of hidden layers of neural network is too large, it will lead to gradient disappearance and gradient explosion.

（5）It is difficult to collect training set samples. It is difficult to collect training set samples used for training neural networks. On the one hand, training neural networks requires a large amount of sample data, and the sample data should be at least one hundred million. On the other hand, the sample data to be collected for text classification needs to keep the same distribution in the training set, the test set and the verification set. This also increases the difficulty of sample data collection to a certain extent.

（6）At present, the accuracy of text classification by neural network algorithm is better than KNN algorithm and SVM algorithm, but the accuracy of text classification by neural network algorithm still exists. The current neural network model can't reach the optimal error of manual classification, and the neural network model still needs further optimization.

（7）Back propagation of neural networks The back propagation algorithm used in neural networks has a certain optimization space [13], and it needs a lot of data for training. The future development trend of text classification by neural network algorithm is to use less data in order to achieve better performance.

1.2 RESEARCH STATUS

In recent years, with the improvement of hardware performance and the increase of data, deep learning has been widely used in computer vision, speech recognition and natural language processing.

Computer vision is the earliest application field of deep learning algorithm, and it is also one of the most active research directions of deep learning application. In the ILS VRC competition in 2012, Krizhevsky et al. trained a large-scale deep convolution neural network Alex Net, which contains eight learning layers: five convolution layers and three full connection layers. In the end, it won the title of ILS VRC competition classification project in 2012 with the error rate of Top-5 of 15.3%, which also made deep learning a hot spot in research. The model designed by Zeiler et al. in

2013 reduced the error rate to 11.2%. Google Net designed by Christianszegedy et al. In 2014 is a 22-layer convolutional neural network, and the error rate of Top-5 in ILS VRC 2014 is 6.7%. In the same year, the runner-up was VGG model designed by Simonyan et al. The error rate in TOP-5 was 7.3%, which also dropped below 10%. In 2015, Kaiminghe et al. of Microsoft Research Asia designed a ResNet architecture with as many as 152 layers. It is deeper than the previous deep neural network, but it reduces the complexity, thus making it easier to train. The error rate of ResNet has also dropped to 3.6%, while the human level is only 5%-10%. It can be seen that the deep learning network has been comparable to the discrimination ability of human beings. In 2017, the SENET model of Jiehu et al. reduced the TOP-5 error rate to 2.3%. In 2017, Andrewy. Ng et al. applied deep learning technology to medical image recognition. The proposed Chexnet algorithm is a 121-layer convolutional neural network, which can identify patients with pneumonia from X-ray chest radiographs, and its diagnostic accuracy exceeds that of professional radiologists.

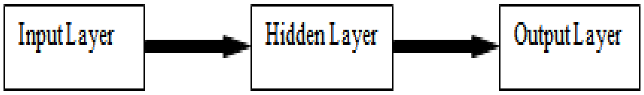
The task of speech recognition is mainly to convert an acoustic signal including natural language pronunciation into the corresponding word sequence. For a long time, the speech recognition system mainly adopts Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) as an acoustic model. Since 2010, the task of using deep neural network for speech recognition has been paid attention to. Microsoft first applied deep learning to its own speech recognition system. In 2011, Microsoft proposed a speech recognition model for large vocabulary based on context-dependent deep neural network-Hidden Markov Model. In the same year, a model T based on context-dependent deep belief network-hidden Markov model was proposed to recognize large vocabulary speech. However, reference uses restricted Boltzmann machine to learn better speech sound wave representation. Literature sparse phoneme recognition with multilayer perceptron. Literature uses deep belief network to model acoustics. In reference, it is proposed to use Deep Belief Network (DBN) for phoneme recognition. Compared with the traditional speech recognition system, its performance is higher. In literature, the method of pruning nodes was proposed to reconstruct DNN, and it was proved to be effective. In reference, the recurrent neural network is used for speech recognition, and its performance is improved. Literature introduces the application of deep learning technology from speech recognition to language and multi-mode processing tasks in industry.

At present, Microsoft, Google and other institutions are carrying out deep neural network research in the field of speech recognition, and have corresponding products applied to our lives, such as Siri of Apple, virtual speech assistants such as Cortana of Microsoft, speech recognition systems of Baidu and Iflytek, etc. In addition, applications such as Voice Search (VS) and Short Message Dictation (SMD) all use the latest speech recognition technology.

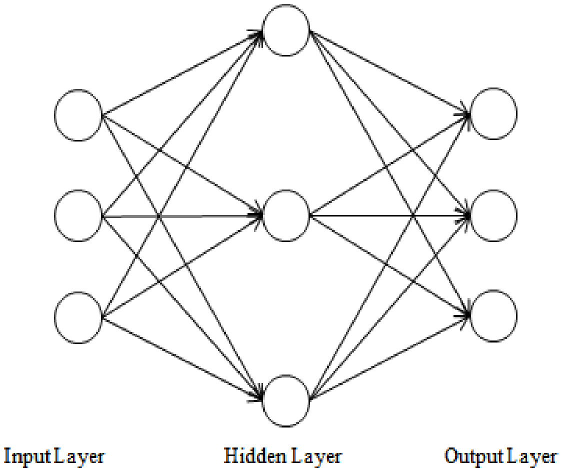
2. INTRODUCTION OF RELATED TECHNOLOGIES AND THEORIES

2.1 INTRODUCTION OF NEURAL NETWORKS

Artificial neural network is a research hotspot in the field of artificial intelligence in recent years, including supervised learning and unsupervised learning. In this paper, supervised learning is used for text classification. By studying the structure principle of biological neural network, supervised neural network constructs a certain functional relationship between a large number of data and label values through training, and then the established model is applied to the same distributed test set data. ANN is composed of a large number of neurons, each of which is composed of linear weighting and activation function. The weight and offset of each neuron are adjusted by a large number of training data to approximate the optimal result as much as possible, so as to process information. The basic structure of neural network is shown in [Fig.2]. The input layer represents the feature input of relevant data after feature extraction, the output layer is the predicted output label value of neural network model, and the hidden layer is the key point of a neural network structure. The hidden layer generally contains more than one layer of network structure. Nowadays, the research of neural network has proved that in most cases, the deep neural network structure can show better performance than the shallow neural network structure, so the hidden layer contains multiple layers of neural networks, and each layer of neural network is composed of multiple neurons.

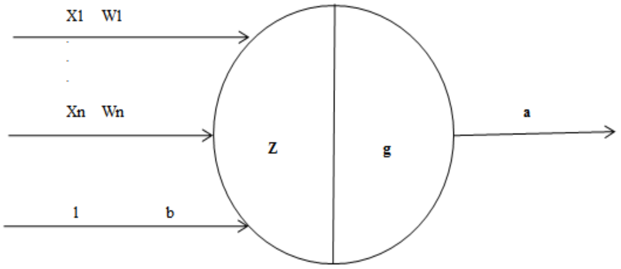


[**Fig.1**] Neural Network Structure Flow Diagram



[**Fig.2**] Structure diagram of neural network

[Fig.2] shows a neural network structure with only one hidden layer, which is a traditional neural network structure, also known as Feedforward Neural Network (FNN). In FNN structure, the nodes of each layer, that is, neurons, are connected with all neurons in the next layer, and the output of each layer of neural network is only related to the input of that layer of neural network, so there is no feedback connection between the output of neural network and the model itself. The output number of each neural network is the same as the number of neurons in this neural network. Generally speaking, the number of layers of neural network does not include the input layer. Neural network is composed of a large number of neurons, just like the neural network of human brain. The structure of each neuron in FNN is shown in [Fig.2].



[**Fig.3]** Neuron structure

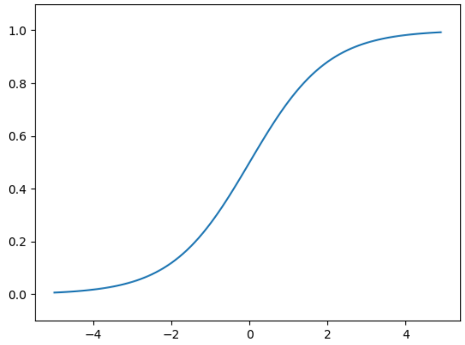
As shown in the [Fig.3], $x_1, x_2, \cdots, x_n$ is the input of the input layer or the output of the upper neural network, etc., $w_1, w_2, \cdots, w_n$ indicating the corresponding weight, and bis the offset.

$$z = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + b \tag{1}$$

Z represents the cache of neurons, and z is the sum of all the input values multiplied by the corresponding weights plus the offset B.

$$\alpha = g(z) \tag{2}$$

Equation 2 indicates that the output value a is $g(Z)$, indicating the activation function, and the output value a is obtained by nonlinear processing of the buffer value z. The commonly used activation functions include sigmoid function, tanh function, Re LU function, etc.

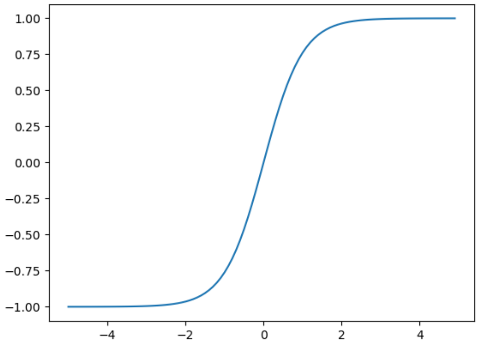


[**Fig.4**] Sigmoid function

Sigmoid function is expressed as $\partial(x) = \dfrac{1}{1+e^{-x}}$ an activation function commonly used in deep learning, and its image is shown in Figure 4. When the value of x is too large, the sigmoid function is close to 1. Similarly, when the value of x is too small, the sigmoid function is close to 0. In both cases, the derivative of sigmoid function will be close to 0. When the value of x is 0, sigmoid function is 0.5, and the reciprocal reaches the maximum value. However, sigmoid function as an activation function has the following two disadvantages.

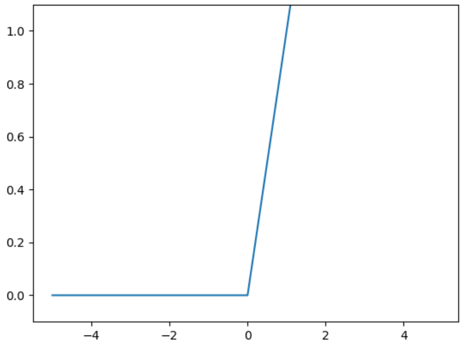
（1） When the value of z is extremely large or small, the derivative of sigmoid function is close to 0. Therefore, in the back propagation, the derivative of weight w will also approach 0, which will lead to the gradient being too small or even disappearing.

（2） In the calculation of neural network, we often need to normalize the data so that the neural network can approach the optimal solution more quickly. However, the output of sigmoid function is not averaged by 0. Therefore, sigmoid function is generally used in the last layer of the network instead of the hidden layer.



[**Fig.5**] Hyperbolic tangent function

[Fig.5] shows the hyperbolic tangent function, $\tanh(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$,which is expressed as the hyperbolic tangent function (tanh) which is very close to the activation function sigmoid. The main difference between Sigmoid function and tanh function is that the former has a range of (0,1) and the latter has a range of (-1,1). Compared with sigmoid function, the average output value of tanh function is 0, but there is also a gradient problem. Based on this problem, the Re LU function shows better performance.

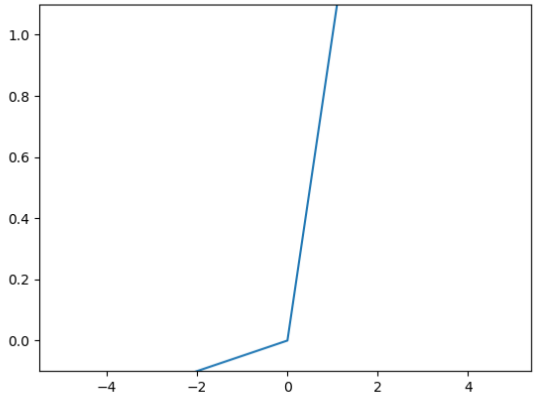


[**Fig.6**]ReLu function

Lu function, also known as modified linear unit, as shown in [Fig.6], is a piecewise function, which solves the gradient disappearance problem of sigmoid function and tanh function. The expression of re function is equation 3.

$$g(Z)=\begin{cases} z, z>0 \\ 0, z<0 \end{cases} \tag{3}$$

When the input is positive, the derivative of Re LU function is 1, so there is no gradient disappearance problem. Lu function is faster than sigmoid function and tanh function in both forward and backward propagation. However, when the input is negative, the gradient of Re LU function is 0, which will lead to the problem of gradient disappearance. Therefore, a kind of Leaky Re LU function appears, and the function image is shown in [Fig.7].



[**Fig.7**]PRe LU function

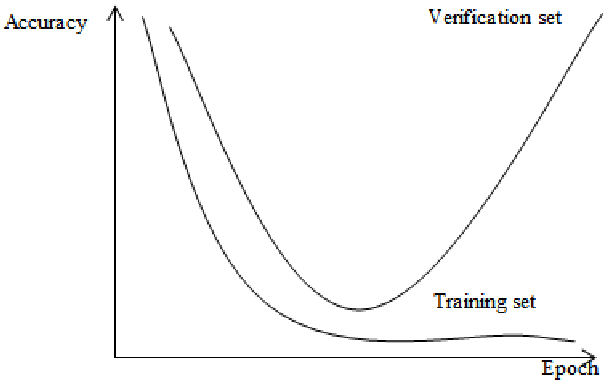
The pre-Lu function is shown in Formula 2-4, which solves the problem that the gradient of re-Lu function disappears when the input is negative.

$$g(Z)=\begin{cases} z,\ z>0 \\ az, z<0 \end{cases} \tag{4}$$

2.2 OVER-FITTING AND UNDER-FITTING PROBLEMS

Under-fitting and over-fitting are common problems in training neural networks. An optimal neural network model should have neither under-fitting nor over-fitting. When training neural network models, it is generally necessary to divide the data samples to be used into training sets, verification sets and test sets according to a certain proportion. The training sets are used to train neural network models, the verification sets are used to intuitively reflect the quality of a model's training in the training process, and the test sets are used to finally test the related performance of a neural network model. The training sets, verification sets and test sets need to keep the same distribution [13]. The neural network model of under-fitting training can't accurately fit the corresponding mapping function, and the network model is too simple, showing high deviation and low variance in training set and verification set [14].

Over-fitting means that the neural network has over-fitted the data such as unnecessary noise, or that the network model is too complex and its fitting ability is too strong, showing low deviation and high variance in the training set and the verification set. The optimal network model should just fit the required mapping function with low deviation and variance. Solving the problem of under-fitting in training neural network can generally increase the number of neurons, increase the number of features of input data or build a deeper neural network model, and try to reduce the deviation by switching to other algorithms [15-19]. The problem of over-fitting caused by too simple data samples can be solved by increasing the sample data. However, the data samples with the same distribution are difficult to collect and can be solved by changing the current sample data. For example, in image samples, the number of samples can be increased by vertical mirror symmetry, random trimming of images, local bending of images, changing resolution, increasing noise and adding different distortion values to RGB three channels. Contrary to the under-fitting problem, the over-fitting problem caused by the complexity of the network model can be solved by reducing the number of neurons, the depth of the network, the number of features of the input data, stopping training in advance and regularization [20].



[**Fig.8**] Error rate chart on training set and verification set

As shown in Figure 8, with the increase of training times, the error rate on the training set, that is, the cost function, decreases gradually, while the error on the verification set decreases first and then increases. On the left side of the best point of the validation set, there is under-fitting and on the right side there is over-fitting. When the validation set reaches the best point, it means that the network model has neither under-fitting nor over-fitting. Therefore, when the network model reaches the best point on the validation set, the training can be stopped in advance to prevent over-fitting.

3. INTRODUCTION TO TEXT CLASSIFICATION

3.1 INTRODUCTION OF DATA SET

The data set used for text classification is generally called corpus, and the expected corpus can be divided into balanced corpus and unbalanced corpus. Balanced corpus means that the number of texts contained in each category is roughly equal, while unbalanced corpus is the opposite. Balanced corpus and unbalanced corpus have extremely important influence on the research of text classification technology. English language database mainly includes 20_Newsgroups data set, Reuters-21578 data set and OHSUMED data set, while Chinese corpus mainly includes Tan Corp V1.0 data set, Sogou Laboratory data set and Fudan University data set. The data set used in this paper is THUC News data set of NLP group in Tsinghua. THUC News data set is processed from the data of Sina RSS subscription channel from 2005 to 2011, and contains 14 categories, including a total of 836075 text data. The specific distribution is as follows:

Finance-economy 37097	**lottery ticket 7588**	**building property 20050**	**Stock 154399**
Education 41940	Science-Technology 162929	Society 50850	Fashion 13368
Current political affairs	Sports 131604	Constellation 3579	Home 32587
63088			
Game 24373	Amusement 92632		

[**Fig. 9**]THUC News data set distribution diagram

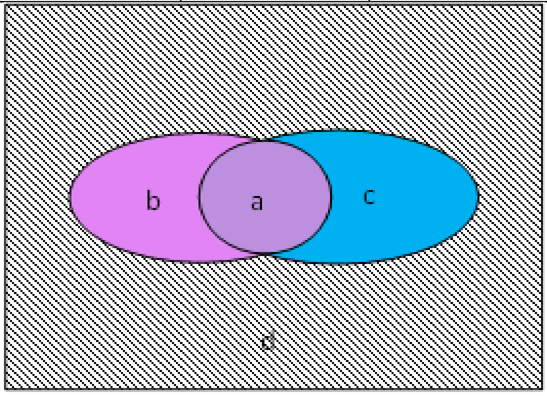
As can be seen from [Fig. 9], THUC News data set is an unbalanced data set, and each text data contains two parts, the title and the body content.

3.2 CLASSIFICATION EVALUATION CRITERIA

Evaluating the performance of a classifier mainly includes Accuracy, Recall, Precision, F1 evaluation value (F1-measure), Macro avg, Micro avg, etc [21-25].

Suppose the test results for a certain category are shown in the following [Fig. 10]:

origin result	T	F
T	a	b
F	c	d



[**Fig.10**] Figure of Classification Test Results

In [Fig. 10], the area A means that the real value is T and the predicted value is T, which is the correct prediction. Area B means that the real value is F and the predicted value is T, which is a false prediction. Area C means that the real value is T and the predicted value is F, which is a false prediction. D refers to the true value of F and the predicted value of F, which is the correct prediction. The formula of recall rate and accuracy rate of a classifier in a certain category is as follows:

$$recall = \frac{a}{a+c} \tag{5}$$

$$precision = \frac{a}{a+b} \tag{6}$$

The evaluation value of F1 can comprehensively analyze the performance of the classifier, as shown in the following formula:

$$F_1 = \frac{2 \times recall \times precision}{recall + precision} \tag{7}$$

The evaluation value of F1 combines recall rate and accuracy rate, which is more accurate for evaluating the performance of classification model. The above classification standards are based on each category, while micro-average is the overall evaluation standard for all categories of the whole classifier. Add A, B, C and D in [Fig.10] according to all categories $a_{all}, b_{all}, c_{all}, d_{all}$, and then calculate the overall recall rate, accuracy rate and F1 evaluation value as above, which is the micro-average. Macros mean the macro average recall rate and macro average accuracy rate according to the recall rate and accuracy rate of each category, and then calculate the macro average F1 evaluation value according to the macro average recall rate and macro average accuracy rate.

In addition, there is the accuracy rate, which is different from the accuracy rate. The accuracy rate is calculated as follows:

$$accuracy = \frac{a_{all} + d_{all}}{a_{all} + b_{all} + c_{all} + d_{all}} \qquad (8)$$

The accuracy rate is the number of all the texts predicted accurately compared with the total number of texts, which directly reflects the performance of a classifier as a whole. There is an Error rate corresponding to the accuracy rate, and the sum of the accuracy rate and the error rate is 1. Confusion Matrix is a kind of matrix that can clearly reflect the actual situation of each category classified by the classifier. It is a matrix with N rows and N columns, and N represents the number of categories. The specific structure is shown in the following figure:

As shown in fig. 11, each row of the confusion matrix [15] adds up to the quantity value of each category, each column adds up to the quantity value predicted by each category, and the left diagonal row is the quantity value predicted by all categories accurately. For example, C21 indicates that the real value is the quantity value predicted by category 2 incorrectly into category 1, C22 indicates that the real value is the quantity value predicted by category 2 correctly into category 2, and C2n indicates that the real value is the quantity value predicted by category 2 incorrectly into category N.

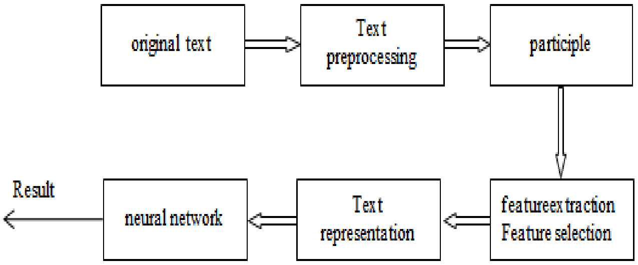
origin result	Categor y1	Categor y2	…	Category n
Categor y1	C_{11}	C_{12}	…	C_{1n}
Categor y2	C_{21}	C_{22}	…	C_{2n}
⋮	⋮	⋮	⋱	⋮
Categor yn	C_{n1}	C_{n2}	…	C_{nn}

[**Fig.11**] Confusion matrix structure

4. RESEARCH ON TEXT CLASSIFICATION BASED ON RNN

4.1 THE PROCESS OF TEXT CLASSIFICATION

Text classification based on RNN network basically includes five steps, and the flow chart is shown in the following [Fig.12]:



[**Fig.12**] Flow chart of text classification

Original text: the collected text information without any processing.

Text preprocessing: the original text is preprocessed to remove unnecessary noise information in the text and become the text to be processed.

Word segmentation: Word segmentation is to segment the text to be processed according to certain rules. Word segmentation is mainly for Chinese texts, and English texts don't need word segmentation. Word segmentation is also a necessary step in Chinese text classification based on word level, while word segmentation is not included in text classification based on character level.

Feature extraction and feature selection: Feature extraction and feature selection are digital representations of segmented words or characters, and then they are converted into vector forms. Generally speaking, word vectors should express the relevant information of words as much as possible, and the vector dimensions should be small to reduce the computation of neural networks.

Text representation: the text to be processed is represented by matrix according to word vector or character vector. Generally, texts of different lengths are filled or cut into texts of the same length.

Neural network: build a suitable neural network for training and save the network model according to the needs. Finally, the classification effect of the network model is comprehensively analyzed through the test set.

4.2 EXPERIMENTAL PROCESS

Through traversing the 65000-text information in the training set, and discarding the punctuation, numbers, special symbols and other information in the text, the number of characters based on character level is 5,819, which only contains Chinese characters and English characters. In addition, because different text lengths are different, it is difficult to train the neural network model. In this paper, all texts are processed in a uniform length, and then all text lengths are averaged, and then all text lengths are processed into average lengths. In this paper, the length of all texts is averaged and rounded to 600. For texts with more than 600 words, only the first 600 words are truncated, and for texts with less than 600 words, padding is added to expand to 600 words. Therefore, it is necessary to add a pad character to expand the text and express words that are not in the vocabulary. Therefore, the vocabulary created in this paper is a 5820-word vocabulary based on character level. Number the created vocabulary according to the vocabulary order, 0 ~ 5819, and then convert it into one-hot vector form, which is a vector that only represents the location attribute. Generally, the dimension of one-hot vector is the number of words in the vocabulary, such as the number of words in (body, education, science, technology, society, Will) and other six words do one-hot vector transformation to ((1 0 0 0 0 0), (0 1 0 0 0 0), (0 0 1 0 0 0), (0 0 0 1 0 0), (00010), (0 0 0 0 0 1)), which is only 1 in the word order, and all other items are 0. Combining the one-hot vectors of all words into a matrix form is a sparse matrix. The vocabulary used in this paper is converted into one-hot matrix form as follows:

$$O_c = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{9}$$

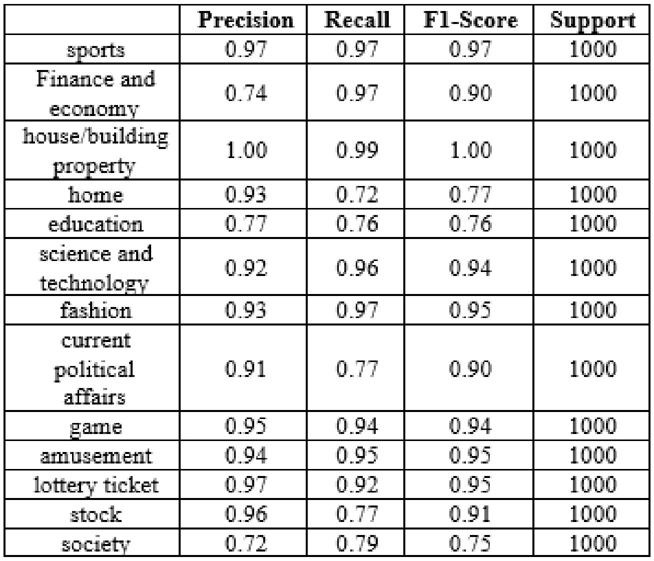
The dimension of the One-hot matrix O_c is (5820, 5820), which mainly has two disadvantages: (1) The dimension of the matrix is too high. If the One-hot matrix is directly used as the input, the number of weight parameters in the neural network will be particularly large, which increases the training difficulty of the neural network and reduces the training convergence speed. (2) The One-hot matrix contains too little information. For example, in Formula 3-1, the one-hot matrix only contains the word order, which can't describe a word comprehensively and accurately, and it doesn't improve the performance of the classifier. Based on the shortcomings of the above two aspects, it is the focus of word vector research to transform One-hot matrix into other matrices with lower dimensions and more information. Dimension reduction of One-hot matrix is called word embedding, and the commonly used embedding method is word2vec software tool developed by Google in 2013. Word2vec is a scheme that uses shallow neural network to train and reduce the dimension of the relationship between words. It mainly includes two models: skip-gram and continuous word bag CBOW. skip-gram is the one-hot vector for inputting specific words and one-hot vector for outputting related words, while continuous word bag CBOW is the opposite of CBOW.

The input of the model is the one-hot vector of a number of related words of a specific word, and the output is the one-hot matrix of a specific word. The number of neurons in the first layer of neural network is the dimension of the embedding word matrix constructed, and the number of neurons in the second layer of neural network is the number V of words in the vocabulary. The parameter matrix W of the first layer of neural network is obtained through many iterations of a large number of text specific words, with the dimension of (N,V). This parameter matrix is the embedded matrix of the trained embedding words, and the column vector is a new vector with dimension N converted from one-hot vector. The dimension of the new vector can be adjusted independently according to the actual needs, and it is much smaller than that of one-hot vector. On the other hand, word2vec vector is trained according to context-related words, which not only contains the information of vocabulary order, but also contains more related information between word meanings.

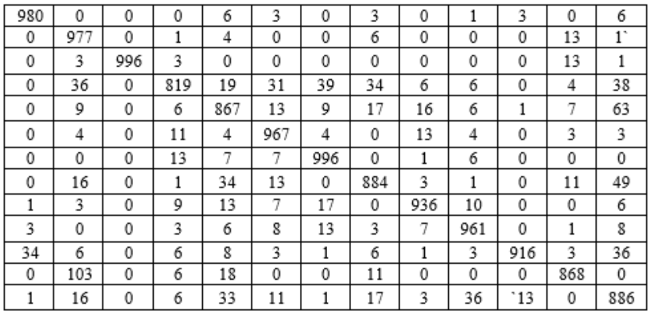
4.3 ANALYSIS OF EXPERIMENTAL RESULTS

The final result of the word vector designed in this paper on the verification set is better than that of word2vec. The precision of the word vector in this paper is 92.02% and the loss function is reduced to 0.33. The accuracy of word2vec word vector on the verification set finally reached 90.85%, and the loss function was reduced to 0.38. Then, the same test set is used for testing, and keep_prob is set to 1, which means that the dropout regularization is not adopted, and the test data in each category are all 1000. The accuracy of word2vec word vector on the test set is 92.23%, and the loss function is 0.28. Micro-average and macro-average are both 0.92, and the indicators such as accuracy and confusion matrix in each category are followed.

[**Table 1]** Word2vec word vector test results



	Precision	Recall	F1-Score	Support
sports	0.97	0.97	0.97	1000
Finance and economy	0.74	0.97	0.90	1000
house/building property	1.00	0.99	1.00	1000
home	0.93	0.72	0.77	1000
education	0.77	0.76	0.76	1000
science and technology	0.92	0.96	0.94	1000
fashion	0.93	0.97	0.95	1000
current political affairs	0.91	0.77	0.90	1000
game	0.95	0.94	0.94	1000
amusement	0.94	0.95	0.95	1000
lottery ticket	0.97	0.92	0.95	1000
stock	0.96	0.77	0.91	1000
society	0.72	0.79	0.75	1000

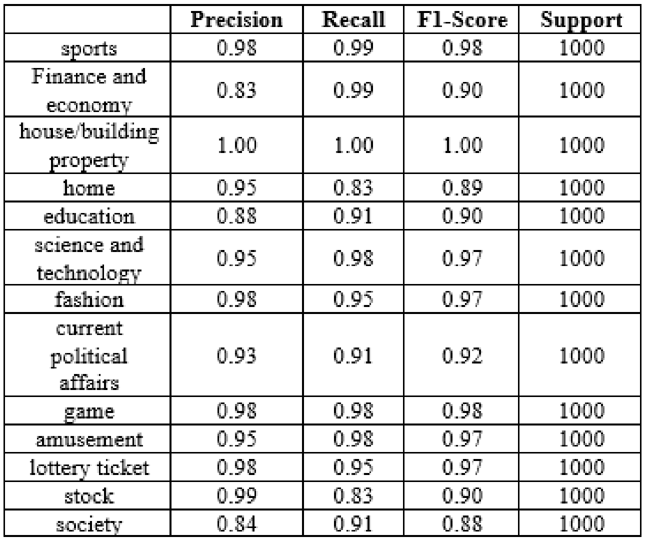


980	0	0	0	6	3	0	3	0	1	3	0	6
0	977	0	1	4	0	0	6	0	0	0	13	1`
0	3	996	3	0	0	0	0	0	0	0	13	1
0	36	0	819	19	31	39	34	6	6	0	4	38
0	9	0	6	867	13	9	17	16	6	1	7	63
0	4	0	11	4	967	4	0	13	4	0	3	3
0	0	0	13	7	7	996	0	1	6	0	0	0
0	16	0	1	34	13	0	884	3	1	0	11	49
1	3	0	9	13	7	17	0	936	10	0	0	6
3	0	0	3	6	8	13	3	7	961	0	1	8
34	6	0	6	8	3	1	6	1	3	916	3	36
0	103	0	6	18	0	0	11	0	0	0	868	0
1	16	0	6	33	11	1	17	3	36	`13	0	886

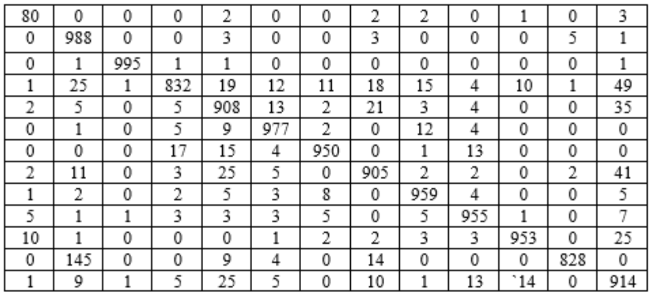
[**Fig.13**] Word2vec word vector confusion matrix

The accuracy of the final test set using word vectors based on word frequency is 93.67%, and the loss function is 0.26. Macro-average and micro-average are both 0.94, which is better than word2vec word vector results. Indicators such as accuracy and confusion matrix in each category.

[**Table 2**] Word Vector Test Results Based on Word Frequency



	Precision	Recall	F1-Score	Support
sports	0.98	0.99	0.98	1000
Finance and economy	0.83	0.99	0.90	1000
house/building property	1.00	1.00	1.00	1000
home	0.95	0.83	0.89	1000
education	0.88	0.91	0.90	1000
science and technology	0.95	0.98	0.97	1000
fashion	0.98	0.95	0.97	1000
current political affairs	0.93	0.91	0.92	1000
game	0.98	0.98	0.98	1000
amusement	0.95	0.98	0.97	1000
lottery ticket	0.98	0.95	0.97	1000
stock	0.99	0.83	0.90	1000
society	0.84	0.91	0.88	1000



80	0	0	0	2	0	0	2	2	0	1	0	3
0	988	0	0	3	0	0	3	0	0	0	5	1
0	1	995	1	1	0	0	0	0	0	0	0	1
1	25	1	832	19	12	11	18	15	4	10	1	49
2	5	0	5	908	13	2	21	3	4	0	0	35
0	1	0	5	9	977	2	0	12	4	0	0	0
0	0	0	17	15	4	950	0	1	13	0	0	0
2	11	0	3	25	5	0	905	2	2	0	2	41
1	2	0	2	5	3	8	0	959	4	0	0	5
5	1	1	3	3	3	5	0	5	955	1	0	7
10	1	0	0	0	1	2	2	3	3	953	0	25
0	145	0	0	9	4	0	14	0	0	0	828	0
1	9	1	5	25	5	0	10	1	13	`14	0	914

[**Fig.14]** Word vector confusion matrix Based on Word Frequency

As can be seen from the above figure, word vector based on word frequency is superior to word2vec word vector in most categories after testing. For example, in the category of real estate, word vector based on word frequency shows excellent performance, with the accuracy, recall rate and F1 evaluation value all being 1, while the recall rate of word2vec word vector in the category of real estate is 0.99. Through the confusion matrix in the above figure, we can see that the classification effect of word vectors based on word frequency in 11 categories, such as sports, finance, real estate, etc., is better than that of word2vec, but only in fashion and stock, the classification effect is weaker than that of word2vec.

5. CONCLUSION

With the development of Internet, people are exposed to more and more data information in daily life, so it is an urgent technical problem to deal with related information more quickly and accurately. Using neural network to deal with Chinese text classification has better classification effect than traditional manual classification, naive Bayes method, decision tree method, etc. However, Chinese text classification technology is completely different from English text classification technology because of its own characteristics, and there are word segmentation methods and feature selection. There are many problems in Chinese text classification, such as word embedding matrix, etc. In addition, there are still many problems in Chinese text classification, such as long training time, weak generalization ability of network model, under-fitting and over-fitting, gradient disappearance and gradient explosion. In this paper, the cyclic neural network and convolutional neural network are respectively studied and optimized in Chinese text classification technology. The specific work is as follows: Aiming at the problem that word2vec word vector needs extra training and the dimension meaning is difficult to distinguish, this paper designs a word vector based on word frequency for classification. Through experimental comparison, it is found that the word vector based on word frequency designed in this paper is superior to word2vec word vector in training convergence speed and final classification effect. This chapter first introduces the general steps of using neural network to classify Chinese texts, then preprocesses the original texts used in the experiment, and then designs a word vector suitable for Chinese text classification based on word frequency, and compares it with word2vec word vector through the experiment. The results show that the word vector designed in this paper is better than word2vec word vector in training convergence speed, accuracy of verification set and classification results. Finally, this paper designs a network model based on Bi LSTM structure, which is composed of three-layer tower-type circulating neural network and deep neural network. Experiments show that the network model designed in this paper is superior to Text CNN network model in training convergence speed and final classification results.

ACKNOWLEDGEMENT

This work was supported by Dongseo University, "Dongseo Cluster Project" Research Fund of 2022 (DSU-20220005)

DECLARATIONS

Author declare that all works are original and this manuscript has not been published in any other journal.

REFERENCES

[1] K. He, X. Zhang, S. Ren, Deep Residual Learning For Image Recognition[C]//, Proceedings of the IEEE conference on computer vision and pattern recognition, (2016), pp.770-778. https:// doi.org/ 10. 1109/ CVPR.2016.90

[2] G. S. Jie Hu, L. Shen, Squeeze-and-Excitation Networks[J]. Preprint arXiv: 1709.01507, (2017).

[3] P. Rajpurkar, J. Irvin, K. Zhu, CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning[J], (2017).

[4] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning[M]. MIT press. (2016).

[5] C. H. Yu, S. P. Wang, J. J. Guo, Learning Chinese Word Segmentation Based On Bidirectional GRU-CRF and CNN Network Model[J]. International Journal of Technology and Human Interaction (IJTHI), (2019), Vol.15, No.3. https://doi.org/ 10. 4018/IJTHI.2019070104

[6] Daehyon Kim, "Deep Learning Neural Networks for Automatic Vehicle Incident Detection", Asia-pacific Journal of Convergent Research Interchange, SoCoRI, ISSN : 2508-9080 (Print); 2671-5325 (Online), Vol.4, No.3, September (2018), pp. 107-117, https://doi.org/10.14257/apjcri.2018.09.11

[7] Daehyon Kim, "Application of Deep Neural Network Model for Automated Intelligent Excavator", Asia-pacific Journal of Convergent Research Interchange, FuCoS, ISSN : 2508-9080 (Print); 2671-5325 (Online), Vol.6, No.4, April (2020), pp. 13-22, http://dx.doi.org/10.21742/apjcri.2020.04.02

[8] Won-seok Bang, Sun-Hwa Kim, Kuk-Hoan Wee, "A Study on the Effect of CSV on the Performance of Partner Companies: An Analysis of Artificial Neural Networks", Asia-pacific Journal of Convergent Research Interchange, FuCoS, ISSN : 2508-9080 (Print); 2671-5325 (Online), Vol.6, No.7, July (2020), pp. 125-132, http://dx.doi.org/ 10.47116/ apjcri.2020.07.12

[9] Daehyon Kim, "Nonlinear Normalization Model to Improve the Performance of Neural Networks", Asia-pacific Journal of Convergent Research Interchange, FuCoS, ISSN : 2508-9080 (Print); 2671-5325 (Online), Vol.6, No.11, November (2020), pp. 183-192, http://dx.doi.org/10.47116/apjcri.2020.11.16

[10] Sung-Bong Jang, "Deep Neural Network Structure Design for Equipment Failure Prediction in Smart Factory", Asia-pacific Journal of Convergent Research Interchange, FuCoS, ISSN : 2508-9080 (Print); 2671-5325 (Online), Vol.7, No.12, December (2021), pp. 1-10, http://dx.doi.org/ 10.47116/ apjcri.2021.12.01

[11] S. L. Darshan, C. D. Jaidhar, Performance evaluation Of Filter-Based Feature Selection Techniques In Classifying Portable Executable Files[J]. Procedia Computer Science, (2018), Vol.125, pp.346-356. https://doi.org/10.1016/j.procs.2017.12.046

[12] E. Hancer, Differential Evolution For Feature Selection: A Fuzzy Wrapper-Filter Approach[J]. Soft Computing, (2019), Vol.23, No.13, pp.5233-5248. https://doi.org/10.1007/s00500-018-3545-7

[13] S. Maldonado, J. López, Dealing With High-Dimensional Class-Imbalanced Datasets: Embedded Feature Selection for SVM Classification[J]. Applied Soft Computing, (2018), Vol.2018, No.67, pp.228-246. https://doi.org/10.1016/j.asoc.2018.02.051

[14] M. Y. Zhang, X. B. Ai, Y. Z. Hu, Chinese Text Classification System On Regulatory Information Based on SVM[C]//IOP Conference Series: Earth and Environmental Science. (2019), pp.252. https://doi.org/10.1088/1755-1315/252/2/022133

[15] H. M. LI, H. N. Huang, X. Cao, Falcon: A Novel Chinese Short Text Classification Method[J].Journal of Computer and Communications, (2018), Vol.6, pp.216-226. https://doi.org/10.4236/jcc.2018.611021

[16] Jae Yoon Lee, Mounika Durbha, "Customary Broadcast Encryption with Advanced Encryption and Short ciphertexts", Asia-pacific Journal of Convergent Research Interchange, SoCoRI, ISSN : 2508-9080 (Print); 2671-5325 (Online), Vol.2, No.2, June (2016), pp. 27-33, http://dx.doi.org/ 10.21742/ APJCRI.2016.06.04

[17] T. Sai Raaga Sowmya, "Cost Minimization for Big Data Processing in Geo-Distributed Data Centres", Asia-pacific Journal of Convergent Research Interchange, SoCoRI, ISSN : 2508-9080 (Print); 2671-5325 (Online), Vol.2, No.4, December (2016), pp. 33-41, http://dx.doi.org/ 10.21742/ APJCRI. 2016.12.05

[18] Sung Kyu Yun, Nan Young Kim, Hyung-ji Chang, "How to Develop the Information and Knowledge Processing Competency of College Students of Humanities ", Asia-pacific Journal of Convergent Research Interchange, FuCoS, ISSN : 2508-9080 (Print); 2671-5325 (Online), Vol.7, No.5, May (2021), pp. 21-30, http://dx.doi.org/ 10.47116/ apjcri.2021.05.03

[19] Vikas Trikha, Jinan Fiaidhi, Sabah Mohammed, "Identifying EEG Binary Limb Motor Imagery Movements using Thick Data Analytics", Asia-pacific Journal of Convergent Research Interchange, FuCoS, ISSN : 2508-9080 (Print); 2671-5325 (Online), Vol.6, No.9, September (2020), pp. 169-189, http://dx.doi.org/10.47116/apjcri.2020.09.15

[20] Tejas Wadiwala, Jinan Fiaidhi, Sabah Mohammed, "Thick Data Analytics through Ensemble Techniques: Identifying Personalized EEG Biometrics based on Eye State Prediction", Asia-pacific Journal of Convergent Research Interchange,

FuCoS, ISSN : 2508-9080 (Print); 2671-5325 (Online), Vol.6, No.10, October (2020), pp. 41-63, http://dx.doi.org/ 10.47116/apjcri.2020.10.04

[21] G. S. Wang, X. J. Huang, Convolutional Neural Network Text Classification Model Based on Word2vec And Improved TF-IDF[J]. Journal of Chinese Computer Systems, (2019), Vol.40, No.5, pp.1120-1126.

[22] Li Y , Fan B , Zhang W , TireNet: A high recall rate method for practical application of tire defect type classification[J]. Future Generation Computer Systems, (2021), Vol.125, No.3. https://doi.org/ 10. 1016/ j.future.2021.06.009

[23] S. M. Gow, Kellett H A , Toft A D , Accuracy and precision of five analog radioimmunoassays for free thyroxin compared.[J]. Clinical Chemistry, (2020), No.11, pp.11.

[24] Chen P . Hotel Management Evaluation Index System Based on Data Mining and Deep Neural Network[J]. Wireless Communications and Mobile Computing, (2021), Vol.2021, No.2, pp.1-11. https:// doi.org/10.1155/2021/2955756

[25] C. S. Hong, Confusion Plot for The Confusion Matrix[J]. Journal of the Korean Data and Information Science Society, (2021), Vol.32, No.2, pp.427-437. https:// doi.org/10.7465/ jkdi.2021 .32.2. 427

[26] S. V. Kamble, V. C. Sakhare Machine Learning Application in Textile :An Overview, Textiles Trends, Volume 65, No 07, October 2022, pp 50-52.