ZKG INTERNATIONAL

ISSN: 2366-1313

Enhancing Text Classification: Improved Deep Learning Models

[1] Abdul Rais Abdul Waheed,[2] Nayeemmohammed,[3] Syed Abrar Hussain,[4] Mohammed Ali Bilal

[1]Assistant professor, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd.

abdulrais@lords.ac.in

[2,3,4] BE Student, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd.

barkathbr@gmail.com, syedhussaiin0@gmail.com, bilaltha8@gmail.com

Abstract: *Deep mastering technology is developing rapidly. Convolution neural network (CNN), as the main device of deep studying, has been involved and worried by way of many researchers and has been extensively used in records retrieval, category, facts management, mining and different activities. In order to gain the nearby features and key factors of the textual content, the TCNN-DAM model is proposed by way of mastering the TCNN version, which goals to maximize the consequences of the textual content, enhance the effects of the distribution of the textual content and sell the textual content. Model to better distribute within the Segos news corpus. Tests show that the improved model has appropriate results for distribution, which can be right*

Keywords: text classification; convolution neural network; attention mechanism; key words.

I INTRODUCTION

The text in particular is going thru three ideas, which consist of the description of characteristic vectors [1], function extraction [2], and the usage of elegance algorithms [3]. Although the compound sentence can remedy the hassle of the textual content, whilst the genuine scale is a bit big, it will take time to come to be sturdy and the fact will fall.

Deep technique, the important thing moment in current years, has played the leading function in textual content type, taking advantage of the benefits of wise statistics processing, widely utilized in facts mining, professional mapping, facts type, l sentiment evaluation and plenty of one of a kind areas. . According to the simulation mechanisms of the human mind for the hidden device to understand patterns, this could represent a low degree inside the antique elegance, and the resources of information mining, to create

ISSN: 2366-1313

more articulation in the society of extracted assets[4]. Mikonos et al. Create Word2vec phrase vectors, which cannot control the size of feature vectors, but additionally controls the connection among elements through ignoring position data factors [5]. CNN can self-pick out sources close to the data to improve the mode effect, however it can't attain better functions [6]. Based on this, Kim proposed the convolution neural community model TCNN, which uses a massive quantity of convolution operations to solve the modern-day shortcomings of CNN [7]. In this paper, we particularly use the cross-gram model in word2vec, and on the identical time, DAM is transferred to the TCNN version, so that we will acquire essential records close to the textual content and acquire the excellent content material of the text. Text and improve the relevant content of the textual content.

II IMPROVED DEEP LEARNING MODEL TEXT CLASSIFICATION ALGORITHM

A. Development of the TCNN model

The emergence of the TCNN model has made the use of neural networks in textual content to be successful. The special function of the TCNN version is to obtain the local beauty information in the extraction process by the operation of the match. The downside is the inability to keep things close.

Compared to the TCNN-DAM version and the TCNN version, the advantage is that it can explain many social characteristics of the text, fix the important points, and make the most impact. Kind of thing. The details of the TCNN-DAM model are:

Text files are stored as columns. When making the list, select convolution kernels with a specific size. The intercept length set for each line is 2 cents, so the convolution kernel length set to 3 to 5 is the first class.

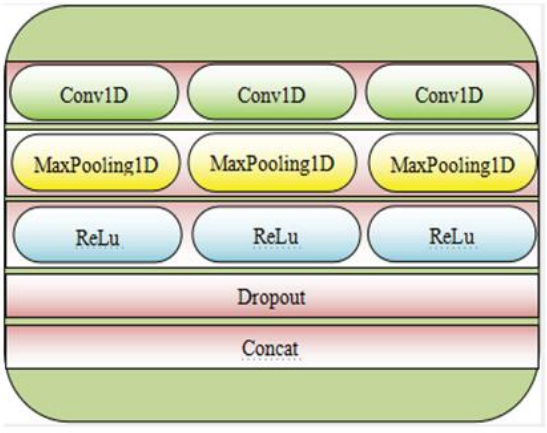

Figure 1 The TCNN model network structure

(2) The dual-channel DAM monitoring mechanism is added to the TCNN model

ISSN: 2366-1313

to give weight to specific data extraction, promote important content, and demonstrate text accuracy.

(3) The output received the Ad max gradient descent process and combined with Soft max for normalization, aiming to reduce the additional output change process.

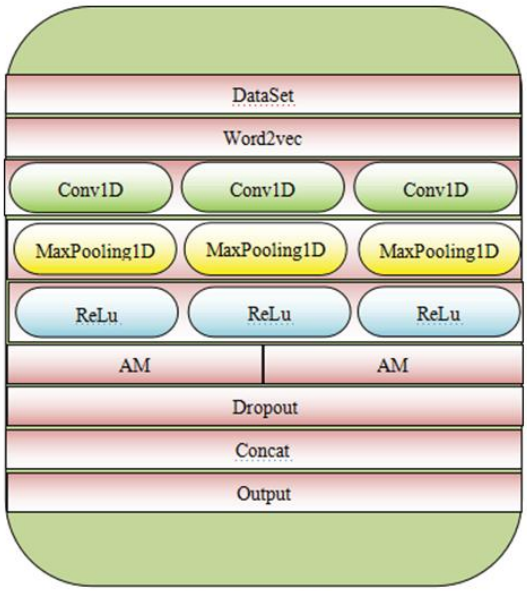

Figure 2 The TCNN-DAM model network structures

c. TCNN-DAM algorithm description

Table 1 the description of the TCNN-DAM algorithm

Algorithm: Based on TCNN-DAM text classification
Input: Sogou News Data Set T **Output:** the category label of the data set to be tested, Accuracy rate Acc, accuracy P, recall rate R and value F_1
(1) Perform word segmentation and remove punctuation on the original text data set D, Word vector initialization and other processing.
(2) The word vector form data.bin file is transported to the convolution layer, and three convolution kernels of different sizes are used, namely 3*128, 4*128, and 5*128, to extract local features of multi-dimensional text.
(3) On the basis of (2), introduce the DAM dual-channel attention mechanism to obtain the TCNN weight information Attention_weights, and highlight the symbolic features.
(4) On the basis of (2) and (3), the multi-dimensional Local features obtained by the convolution layer and the Keyword information obtained by DAM are used to extract the optimal feature information inside using Max Pooling.
(5) Use Dense full connection to send the optimal feature information obtained in (4) to the Soft max classifier, and fully consider the feature classification.

III EXPERIMENT AND ANALYSIS

(1) The setting of the test site immediately determines whether the test can be performed effectively. Setting up the hardware and software environment is very important.

Table 2 Software environment configuration

Software environment	**operating system** Ubuntu14.04	**Development language** Python3.6	**Programming Tools** Jupiter notebook
	Word Vector Tool Word2vec	**Word segmentation tool** Jieba 0.39	**Deep learning framework** Keras+Tensorflow

(2) Test records: This article makes use of the Segou Chinese newspaper corpus. The information changed into accumulated and analyzed by way of Sogou Lab. In this test, 52,000 points have been used. The corpus is divided into eleven agencies, along with

ISSN: 2366-1313

1-QC (automotive), 2-CJ (finance), 3-IT (information technology), four- JK (fitness), five-TY (sports), 6-LY (tourism), 7-JY (education), eight-JS (military), nine-WH (lifestyle), 10-YL (entertainment)), 11-SS (style). After corpus pre-processing, eighty% (41600 rows) of the information corpora are freely blended in line with the training system (schooling data), 20% (10,400 rows) of the information corpus are considered that is the take a look at manner (curriculum). An intuitive picture of the distribution of different forms of corpora is presented in Figure 3.

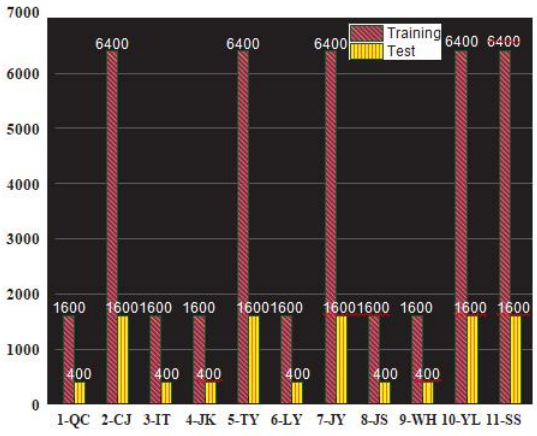

Figure 3 Corpus distribution map of differenttypes of text

B. Test content and assessment

The effectiveness of the method is demonstrated by comparing the tests of the two methods. Experiment Details: (1) Use the Cross-Gram model in Word2vec to introduce the physical first to create 2 hundred-dimensional word vectors, paving the way for the next test. (2) Evaluate the overall classification performance and overall processing time of the TCNN model and TCNN-DAM model on the Sogou data corpus.

Benchmarks measure undeleted performance of the system. In the text category, this includes the version to determine whether the opinion of the text label is correct. Test the model by accuracy, remember, F1 value, accuracy and running time below average macro and average weight.

(1) **The truth**

In fact, it is the percentage of good written content compared to the total recorded data, the real value of the calculation is:

$$F_1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (1)$$

Weights normally consist of WAP, WAR, and WAF, which constitute accuracy, remember, and cost-weighted common, respectively. This is the wide variety of values in line with row in the statistics set, then divided through the range of values in line with row and its anticipated fee. Generally better than predicted on the average macro fee.

(3) **Success**

The requirements that a super custom wishes to be excessive performance and coffee station. Efficiency typically refers to the running time of the algorithm. The test on this paper specially makes use of the new release time of 10 times of one of

ISSN: 2366-1313

the education techniques because the benchmark.

C. **Testing**

The choice of tests plays an essential function in gaining deep understanding and is crucial. LUB

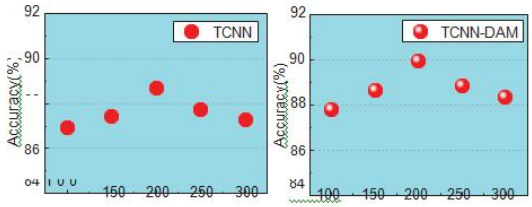

Figure 4 TCNN and TCNN-DAM word vector dimension

(2) In the school system of information content corpus, the cross-validation approach confirmed that the perfect parameters of CNN are: multiple filters 128, convolution kernel size three, 4 and five, optimizer is defined. For Ada max and L2 learning value zero.001, the iteration range is 10.

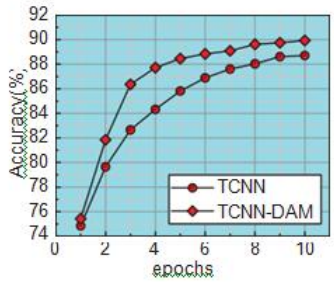

Figure 5 Model performances of TCNN and TCNN-DAM

In this paper, under the same Sogou corpus, after repeated experimental research, the accuracy and time of the improved TCNN model and the improved TCNN-DAM model are summarized in Table 3.

Table 3 Accuracy and time of the two models

Model	Accuracy (%)	Training time(s)
TCNN	88.71	9.5
TCNN-DAM	89.93	11.7

According to Table four, it could be visible intuitively that the accuracy of the TCNN-DAM model is substantially better than that of the TCNN version. The main reason is that DAM is introduced inside the TCNN-DAM model. It is used to fully reap the neighbourhood feature as and keyword data of the text. Under the idea, it may be in addition integrated into the schooling manner of the network. Comparing the schooling time, it is found that the education time of the TCNN-DAM model is delayed in comparison to the TCNN version. The primary reason is that the DAM is introduced, and the community structure of the version turns into extra complex, resulting in a slightly longer education time. But looking at the overall situation, the type impact of the TCNN-DAM model is higher than that of the TCNN model.

V CONCLUSION

This paper presents previous experiments and theories that specialize in using deep learning methods for newspaper classification. Since the TCNN model cannot get the records of textual content and key factors, the TCNN-DAM model is

ISSN: 2366-1313

proposed, so that its precision, recovery charge and cost are better as compared with the preceding model. Of direction, the statistics received in this paper is specific, and more records may be used for training later. At the equal time, research primarily based on textual content class the use of deep mastering may be reinforced.

REFERENCES

1. Samant S, Murthy N L B, Malapati A. Improving Term Weighting Schemes for Short Text Classification in Vector Space Model [J]. IEEE Access, 2019, PP (99).

2. He W, Zhang Y, Yu S, et al. Deep feature weighting with a novel information gain for naive bayes text classification [J]. Journal of Information Hiding and Multimedia Signal Processing, 2019, 10(1):102- 109.

3. Pham B T, Bui D T, Prakash I, et al. Evaluation of predictive ability of support vector machines and naive Bayes trees methods for spatial prediction of landslides in Uttarakh and state (India) using GIS [J]. Journal of Geometrics, 2016, 10(1):71-79.

4. Lin L, Guo S X. Text Classification Feature Extraction Method Based on Deep Learning for Unbalanced Data Sets[M]// Advanced Hybrid Information Processing. 2021.

5. Deep Convolution Neural Network using Stochastic Computing [J]. ACM SIGARCH Computer Architecture News, 2017, 45(1):405-418.

6. Kim Y. Convolution neural networks for sentence classification [C]//Empirical Methods in Natural Language Processing. 2014:1746- 1751.

7. Prasadi Peddi and Dr. Akash Saxena (2015), "The Adoption of a Big Data and Extensive Multi-Labled Gradient Boosting System for Student Activity Analysis", International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 3, Issue 7, pp:68-73.

8. Prasadu Peddi (2016), Comparative study on cloud optimized resource and prediction using machine learning algorithm, ISSN: 2455-6300, volume 1, issue 3, pp: 88-94.