EAI Endorsed Transactions
on Scalable Information Systems

Research Article EAI.EU

Multi-attention mechanism based on gate recurrent unit for English text classification

Haiying Liu[1,*]

[1]School of Foreign Languages, Zhengzhou University of Science and Technology, Zhengzhou 450064 China

Abstract

Text classification is one of the core tasks in the field of natural language processing. Aiming at the advantages and disadvantages of current deep learning-based English text classification methods in long text classification, this paper proposes an English text classification model, which introduces multi-attention mechanism based on gate recurrent unit (GRU) to focus on important parts of English text. Firstly, sentences and documents are encoded according to the hierarchical structure of English documents. Second, it uses the attention mechanism separately at each level. On the basis of the global object vector, the maximum pooling is used to extract the specific object vector of sentence, so that the encoded document vector has more obvious category features and can pay more attention to the most distinctive semantic features of each English text. Finally, documents are classified according to the constructed English document representation. Experimental results on public data sets show that this model has better classification performance for long English texts with hierarchical structure.

Keywords: English text classification, multi-attention mechanism, GRU, deep learning.

Received on 11 January 2022, accepted on 19 January 2022, published on 27 January 2022

doi: 10.4108/eai.27-1-2022.173160

*Corresponding author. Email: snowberry@qq.com

1. Introduction

Text classification is one of the core tasks of natural language processing [1-3]. It is widely used in news classification, sentiment analysis, subject detection and other fields. Traditional machine learning classification algorithms use feature engineering to select suitable features to represent text, and then input text features into different classification models to obtain classification results, such as, Naive Bayes (NB) [4], K-nearest neighbor (KNN) [5], Support Vector Machine (SVM) [6] and other algorithms. However, with the increasing number of texts, the complex feature engineering in the traditional machine learning classification algorithm has become a bottleneck restricting its development.

Deep learning models [7-9] can autonomously learn the characteristics of samples from large-scale data samples, which improves the intelligence of modeling and simplifies the classification process. Therefore, it has become a research hotspot in the field of text classification. Convolutional neural Network (CNN) [10,11] is one of the commonly used models in deep learning. Yoon Kim et al. [12] proposed the TextCNN model in the EMNLP conference in 2014. He used convolutional neural network to model and classify texts and obtained results that were not inferior to the complex classifier model based on machine learning. Therefore, the research on text classification model based on deep learning is hot.

In natural language processing tasks, recurrent neural networks (RNN) [13-15] typically process one word at a

time and learn features based on complex sequences of words. Thus, RNN can capture language patterns useful for natural language processing tasks, especially on long text segments. Francisco et al. [16] obtained good results by using the recurrent neural network based on long and short-term memory (LSTM) to encode text.

Convolutional neural network can extract local features and RNN can extract global features, both of them show good results. However, when the document length is long, processing the document directly as a long sequence not only poses a significant performance challenge to the model, but also ignores the information contained in the document hierarchy. Therefore, some researchers study hierarchical neural network model [17-20] and use hierarchical network model for text classification. However, the hierarchical network model usually uses the global object vector in the training process and cannot pay attention to the most obvious semantic features of each text.

When using attention mechanisms for sentences, the existing methods usually use global parameters as target vectors for all categories. On the one hand, this is not conducive to the expression of the characteristics of each sentence, and on the other hand, it cannot highlight the words with obvious categorical characteristics in the sentence. Aiming at the advantages and disadvantages of current text classification algorithms in long text classification, this paper proposes an English text classification model based on multi-layer attention mechanism combining with GRU [21,22] (abbreviated as MAMGRU). A hierarchical model based on RNN is used to classify long English texts, and an attention mechanism is introduced to pay attention to important parts of English texts. At the sentence attention level, besides training a global objective vector, the most important information in each dimension is extracted from the sentence word vector matrix by maximum pooling as the sentence-specific objective vector. The two object vectors are used to score the words in the sentence together, so that the category features of the sentence coding can be more obvious, and the most distinctive semantic features of each English text can be better paid attention to.

Finally, experiments on public data sets verify the validity of the MAMGRU model. Especially for the data set with hierarchical features and long text, the MAMGRU model obtains specific important features of each sentence by extracting the maximum features of each dimension in the sentence at the sentence attention layer, which has a good performance in the accuracy of text classification.

2. Related works

Deep learning-based text classification algorithms usually use low-dimensional and real-value word vectors to represent the words in the text, then build a neural network model to model the text, obtain the text representation containing all the text information, and use the final text representation to classify the text.

TextCNN model is a classical text classification model based on deep learning. After that, researchers put forward many improvement schemes based on TextCNN. Londt et al. [23] proposed a character-level classification model based on CNN, adding a cyclic layer on the basis of CNN to capture the information that sentences rely on for a long time. Jang et al. [24] obtained sequential dependency information by increasing the depth of CNN. Paoletti et al. [25] studied how to deepen the word granularity of CNN to express the text globally, and proposed a simple pyramidal CNN network structure, which not only increased the network depth and accuracy, but also did not increase the amount of calculation too much.

Inspired by TextCNN and considering the advantages of RNN in processing text data, some researchers use RNN and its variant structure to model and classify texts. Shibata et al. [26] used LSTM structure to encode text, and in order to solve the problem of less annotated data in a single task, they designed three different information sharing mechanisms based on RNN for training, and achieved good results in the four benchmark text classification tasks. In order to solve the problem of insufficient storage unit when RNN encoding long text, Liang et al. [27] proposed a LSTM structure with high caching to capture the overall semantic information in long text, so that the network could better preserve emotional information in a circular unit.

In addition to capturing sequence information on long text, another advantage of RNN for text classification is that it can be well combined with the attention mechanism. It can focus on key information in text modeling to improve the effect of English text classification. The attention mechanism is first proposed in the field of computer vision [28,29]. In the field of natural language processing, the attention mechanism is first introduced into the RNN based encoding model of machine translation tasks. The attention mechanism scores the input sequence by object vector, focuses attention on the more important part of the input sequence, and makes the output result more accurate. Therefore, it has been gradually promoted and applied to a variety of NLP tasks including text classification.

When dealing with a long document representation consisting of many sentences, treating the document directly as a long sequence ignores the information contained in the document hierarchy. Therefore, some researchers use hierarchical neural network model to model documents for text classification. Hu et al. [30]

constructed a bottom-up document representation method. CNN was first used to encode sentences, then RNN with gated structure was used to construct document representation, and finally classification results were obtained through Softmax layer. Experiments show that this model achieved the best result for long document classification at that time. Similarly, Ouyang et al. [31] proposed the hierarchical attention model, which incorporated the attention mechanism into the hierarchical GRU model, enabling the model to better capture the important information of documents and further improved the accuracy of classification of long English documents.

3. Proposed English text classification model

To capture the sequence information on the text in English, at the same time, better use of long hierarchy of document data, this paper puts forward the text classification model based on multi-attention mechanism, using the hierarchical model based on RNN, introducing the mechanism of focusing on the most important part of English text, in the words of attention by extracting sentences of each dimension in the biggest characteristics to obtain the importance of each sentence.

The basic idea of the model is to encode sentences and documents respectively according to the hierarchical structure of words forming sentences and sentences forming documents. In order to focus the semantics of a sentence or document representation on its important components, attention mechanisms are used at each level. Finally, the documents are classified according to the document representation built.

Usually the attention mechanism starts by setting up a task-specific goal vector and matching it to the input sequence. Each element in the input sequence is then assigned an attention score by calculating its similarity to the target vector. After the score is normalized, the weighted sum of all elements gives the final representation. In text classification, the previous practice is to learn a global context vector in the network as the object vector, and score words by calculating the similarity between each word and the object vector.

However, when all categories share an object vector, its information on each feature dimension will be relatively average and it cannot highlight the salient features of the sentence. Even if a word with distinct categorical features appears in a sentence, the global object vector cannot assign it an attention score matching its salience.

Therefore, the MAMGRU model uses a mixed attention mechanism at the sentence attention level, that is, in addition to using the global objective vector, each sentence constructs its own specific objective vector. The largest value in each dimension, that is, the most obvious information, is directly extracted from the word vector matrix of the sentence as the specific object vector of the sentence. The MAMGRU model has five layers, including sentence coding layer, sentence attention layer, document coding layer, document attention layer and document classification layer from bottom to top. The structure of the model is shown in Figure 1.

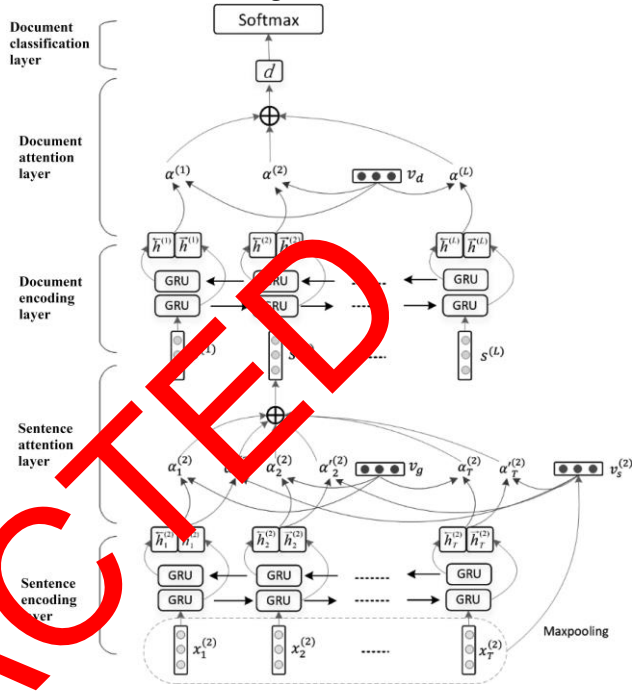

Figure 1. Proposed English text classification model

3.1. Sentence encoding layer

At the sentence encoding layer, each word in the sentence is input in turn to construct the sentence representation. To obtain the sequence information in the sentence, the sentence is modeled using RNN. However, in the basic RNN, as the sequence spreads over time, the historical information in the sequence will be gradually forgotten, while the error accumulation is increasing. Therefore, in MAMGRU, a special RNN structure-gate recurrent unit (GRU) is used to solve the gradient updating problem in long-term memory and back propagation.

For a sentence $s^{(i)} = [x_1^{(i)}, x_2^{(i)}, \cdots, x_T^{(i)}]$, $x_t^{(i)}$ is the vector representation of the t-th word in sentence $s^{(i)}$, $t \in [1, T]$. Bi-directional GRU is used to obtain word annotations combining contextual information by summarizing word information from both directions, as shown in equations (1) and (2).

$$\vec{h}_t^{(i)} = GRU_1(x_t^{(i)}) \quad (1)$$

$$\overleftarrow{h}_t^{(i)} = GRU_2(x_t^{(i)}) \quad (2)$$

Where, GRU_1 and GRU_2 represent GRU structures in two directions, which are encoded from front to back and back to front respectively. For each word $x_t^{(i)}$, it connects the two directions of the hidden layer states $\vec{h}_t^{(i)}$ and $\overleftarrow{h}_t^{(i)}$ as its annotations $h_t^{(i)}$. This annotation represents the entire sentence information centered on $x_t^{(i)}$, as shown in equation (3).

$$h_t^{(i)} = [\vec{h}_t^{(i)}, \overleftarrow{h}_t^{(i)}] \quad (3)$$

3.2. Sentence attention layer

Because each word in the sentence contributes differently to the classification goal. The weight of words that are more important to the classification should be higher during coding. Therefore, at the sentence attention layer, the attention mechanism is used to calculate an attention score for each word in the sentence, and then the vector representation of the sentence is formed through the word and its score.

In particular, at the sentence attention level, MAMGRU model uses a hybrid attention mechanism. In addition to using the global target vector v_g, it also constructs its unique target vector v_s for each sentence. Since each dimension of the word vector represents an attribute information, similar to max-pooling in CNN [32,33], the largest value in each dimension is directly extracted from the word vector matrix of the sentence, that is, the most obvious information, as the unique target vector of the sentence, and the semantic information with obvious categorical characteristics is more highlighted.

For each sentence in the document, $s^{(i)} = [x_1^{(i)}, x_2^{(i)}, \cdots, x_t^{(i)}]$, $x_t^{(i)}$ is the vector representation of the t-th word in sentence $s^{(i)}$, $t \in [1, T]$. The word vector of each word is $x_t^{(i)} = [x_{t1}^{(i)}, x_{t2}^{(i)}, \cdots, x_{tW}^{(i)}]$, W is the phrase vector dimension.

It calculates the specific object vector $v_s^{(i)}$ of the sentence, combines with the global object vector v_g, and adds them together according to certain weights as the attention score of the sentence to get the final coding value of the sentence. As shown in figure 2, when calculating the specific object vector $v_s^{(i)}$ of the sentence, the maximum value on each dimension of all T words is taken as the feature, and then the maximum value on all W dimensions is connected as the specific object vector $v_s^{(i)}$ of the sentence $s^{(i)}$, as shown in formula (4) and (5).

$$v_s^{(i)} = [u_1^{(i)}, u_2^{(i)}, \cdots, u_W^{(i)}] \quad (4)$$

$$u_j^{(i)} = \max(x_{1j}^{(i)}, x_{2j}^{(i)}, \cdots, x_{Tj}^{(i)}), j \in [1, W] \quad (5)$$

Where, $v_s^{(i)}$ is the specific target vector of sentence $s^{(i)}$. $u_j^{(i)}$ is the j-th dimension of $v_s^{(i)}$. $x_{tj}^{(i)}$ is the j-th dimension of the phrase vector of the t-th term in sentence $s^{(i)}$.

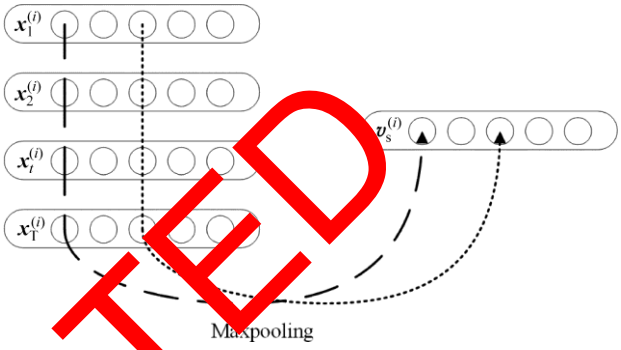

Figure 2. Extracting the target vector by maxpooling

At the same time, a global objective vector v_g is set to represent "which words are more important to the classification objective". It is randomly initialized during training and continuously learned as a parameter.

After obtaining two object vectors, $h_t^{(i)}$ needs to be processed through a layer of fully connected network in order to multiply the word annotation $h_t^{(i)}$ and the object vector, as shown in equation (6).

$$h_t'^{(i)} = \tanh(W_s h_t^{(i)} + b_s) \quad (6)$$

Then, for all the words in the sentence, the similarity between them and the two target vectors is calculated and normalized. The attention score for the two target vectors is obtained, as shown in equations (7)~(10).

$$e_t^{(i)} = a(h_t'^{(i)}, v_g) = \exp(h_t'^{(i)T} v_g) \quad (7)$$

$$\alpha_t^{(i)} = \frac{e_t^{(i)}}{\sum_t e_t^{(i)}} \quad (8)$$

$$e_t'^{(i)} = a(h_t'^{(i)}, v_s^{(i)}) = \exp(h_t'^{(i)T} v_s^{(i)}) \quad (9)$$

$$\alpha_t'^{(i)} = \frac{e_t'^{(i)}}{\sum_t e_t'^{(i)}} \quad (10)$$

Where $h_t'^{(i)}$ is the word representation after fully connected network processing. v_g is the global object

vector obtained by training. $v_s^{(i)}$ is a sentence-specific object vector obtained by maximum pooling on the word vector matrix of sentence $s^{(i)}$. a is the scoring function, and it means the dot plus exponential function. $h_t'^{(i)T}$ is $h_t'^{(i)}$ transpose. $\alpha_t^{(i)}$ and $\alpha_t'^{(i)}$ are the normalized fractions of the two object vectors v_g and $v_s^{(i)}$ corresponding to the word annotation $h_t^{(i)}$ respectively.

It adds the two scores according to a certain weight as the final attention score. The coding of sentences is obtained according to all words and their attention scores, as shown in equation (11).

$$s^{(i)} = \sum_t ((1-\lambda)\alpha_t^{(i)} + \lambda\alpha_t'^{(i)})h_t^{(i)} \quad (11)$$

In this way, the corresponding vector representation can be obtained for each sentence in the document, and the words with obvious classification characteristics in the text will get more weight and occupy a dominant position in the final sentence representation.

3.3. Document encoding and attention layer

After obtaining the vector representation of $s^{(i)}$ of the sentence, bidirectional GRU is also used to encode $s^{(i)}$ in the coding layer of similar sentences, as shown in equation (12) and equation (13).

$$\vec{h}^{(i)} = GRU_1(s^{(i)}), i \in [1, L] \quad (12)$$

$$\overleftarrow{h}^{(i)} = GRU_2(s^{(i)}), i \in [1, L] \quad (13)$$

Then join $\vec{h}^{(i)}$ and $\overleftarrow{h}^{(i)}$ to get the comment $h^{(i)}$ of the sentence $s^{(i)}$. $h^{(i)}$ contains the information of adjacent sentences around $s^{(i)}$ but it still focuses on $s^{(i)}$, as shown in equation (14).

$$h^{(i)} = [\vec{h}^{(i)}, \overleftarrow{h}^{(i)}] \quad (14)$$

Each sentence in the document also contributes differently to the result of the document classification. Therefore, in the document attention layer, the attention mechanism is also used to score each sentence. Here we focus only on the global information of the document. Therefore, a document-level global objective vector v_d is used to measure sentence importance. The target vector v_d represents "which sentences are more important". They are also randomly initialized and co-learned during training.

Similar to the attention layer of sentences, the full connection layer should be used to process the annotation $h^{(i)}$ of sentences, as shown in equation (15).

$$h'^{(i)} = \tanh(W_d h^{(i)} + b_d) \quad (15)$$

Then, the similarity is calculated and normalized according to the global object vector v_d at the document level, and the attention score is obtained. Then, the document vector d containing all sentence information in the document is obtained by means of weighting, as shown in equations (16)~(18).

$$e^{(i)} = a(h'^{(i)}, v_d) = \exp(h'^{(i)T} v_d) \quad (16)$$

$$\alpha^{(i)} = \frac{e^{(i)}}{\sum_t e^{(i)}} \quad (17)$$

$$d = \sum_i \alpha^{(i)} h^{(i)} \quad (18)$$

3.4. Document classification layer

Document vector d is a higher-order representation of document, and can be directly used as a feature of document classification. The probability of each category is calculated by Softmax, as shown in equation (19).

$$p = soft\max(W_c d + b_c) \quad (19)$$

In the whole model training process, the cross entropy is used as the loss function, as shown in equation (20).

$$L = -\sum_d p_{d_j} \log p_{d_j} \quad (20)$$

Where p_{d_j} is the probability that document d belongs to category j.

To sum up, MAMGRU model firstly uses bidirectional GRU and mixed attention mechanism to encode sentences and obtain vector representation of each sentence in the document. The document is then encoded using a two-way GRU and a basic attention mechanism. The resulting document representation contains semantic information for all the sentences in the document, with a greater proportion of important sentences. Similarly, the more distinctive the words in each sentence, the greater the weight. Finally, at the document classification layer, classify documents according to the obtained document representation, and obtain the probability of documents corresponding to each category. Based on such a hierarchy, a more categorical representation of documents can be obtained.

4. Experiments and analysis

In this part, the experiments on a balanced dataset of internet media reports in the Chinese language are carried out. In this dataset, there are 1000 training samples and

1000 test samples in each category. This is an exactly balanced dataset [34].

In the training process, all the data in the data set are divided into training set and test set by 9:1, and 10% of the training set is taken as the verification set.

The experiment compares the MAMGRU model with the following classification models.

(1) Classification model based on machine learning: Bayesian model, decision tree model. The Bayesian model trains Bernoulli Bayesian classifier [35] and polynomial Naive Bayesian classifier respectively. The decision tree model trains the ID3 decision tree based on information entropy [36] and the CART decision tree based on GINI impurity [37].

(2) Classification model based on deep learning: Including TextCNN and TextRNN without hierarchical model, and GRU with hierarchical model [38-40].

Part of training parameter settings of MAMGRU model are shown in Table 1. The size of batch samples is set as 32. The number of hidden units in GRU is 50. The word vector has a dimension of 200. The learning rate of Adam optimization algorithm is 0.001. The threshold of gradient clipping is 5. After adjustment and verification, the hyperparameter input of sentence attention layer achieves the best effect when λ =0.2.

Table 1. Part training parameters setting of MAMGRU

Parameter type	Parameter value
epoch	10
Batch-size	32
Hidden-size	50
Embedding-size	200
Learning-rate	0.001
Grad-clip	5

We compare traditional machine learning algorithm, convolutional neural network text classification model TextCNN, recurrent neural network text classification model and MAMGRU proposed in this paper. The experimental results are shown in Table 2. The ten average classification accuracy of TextCNN model, GRU model and MAMGRU model after the convergence of classification effect is taken as the classification result [41-43].

Table 2. Classification result

Model	Accuracy/%
Polynomial naive Bayes	74.41
Bernoulli Bayes	68.82
KNN	87.46
Information entropy decision tree	87.61
TextCNN	93.29
GRU	94.03
MAMGRU	96.28

Naive Bayes model assumes that attributes are independent of each other, but this assumption is often not true in practical application and the performance of Naive Bayes is poor when the number of attributes is large. When the value of K is 1, KNN algorithm has a better effect on the open English data set. The K value of KNN algorithm has a great relationship with the data set itself. There is no specific empirical formula for determining K value. Therefore, when training the new data set, it is necessary to re-experiment with different values of K to determine the optimal K value. Since information entropy is more sensitive to impurity than Gini coefficient, when information entropy is used as an indicator, the growth of decision tree will be more "fine". When data dimension is high or data noise is high, information entropy is easy to over-fit.

TextCNN has achieved a good classification effect on data sets, which is higher than the accuracy of traditional machine learning classification model. GRU has a slightly better classification effect than TextCNN, and the proposed model achieves the same results as GRU on public data sets. Compared with short texts, the proposed model can pay more attention to the most distinctive semantic features in long texts and achieve better results.

5. Conclusion

This paper studies deep learning based text classification algorithm and proposes a novel English text classification model based on multi-attention mechanism. First, it improves the multi-attention model. In sentence encoding, it is proposed to set a specific objective vector for each sentence in the text, combine with the global objective vector, synthesize all word items and their attention score according to a certain weight to get the sentence coding. Then, through the document coding layer, document attention layer and document classification layer in the

hierarchical attention model, the probability of documents corresponding to each category is obtained, that is, the realization of text classification. The experimental results show that the proposed algorithm is better than the traditional machine learn-based text classification models and the existing deep learn-based text classification models.

References

[1] Mironczuk M M, Protasiewicz J. A Recent Overview of the State-of-the-Art Elements of Text Classification[J]. Expert Systems with Applications, 2018, 106(sep.):36-54.

[2] Labani M, Moradi P, Ahmadizar F, et al. A novel multivariate filter method for feature selection in text classification problems[J]. Engineering Applications of Artificial Intelligence, 2018, 70:25-37.

[3] Lei F, Liu X, Li Z, et al. Multihop Neighbor Information Fusion Graph Convolutional Network for Text Classification[J]. Mathematical Problems in Engineering, 2021, 2021(1):1-9.

[4] Abhilash P M, Chakradhar D. Sustainability improvement of WEDM process by analysing and classifying wire rupture using kernel-based naive Bayes classifier[J]. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 2021, 43(2).

[5] Gao L, Li D, Yao L, et al. Sensor drift fault diagnosis for chiller system using deep recurrent canonical correlation analysis and k-nearest neighbor classifier[J]. [illegible] Transactions, 2021(3).

[6] Yin Shoulin, Liu Jie, Teng Lin. A new kind herd algorithm based on SVM method for road feature extraction[J]. Journal of Information Hiding and Multimedia Signal Processing, v 9, n 4, p 997-1005, July 2018.

[7] Yin Shoulin, Liu Jie, Li Yang. A Self-Supervised Learning Method for Shadow Detection in Remote Sensing Imagery[J]. 3D Research, Vol. 9, No. 4, December 1, 2018. https://doi.org/10.1007/s13319-018-0204-9

[8] Zhao Yue, Li Hang, Yin Shoulin, Sun Yang. A new Chinese word segmentation method based on maximum matching[J]. Journal of Information Hiding and Multimedia Signal Processing, v 9, n 6, p 1528-1535, November 2018.

[9] Shoulin Yin, Ye Zhang and Shahid Karim. Region search based on hybrid convolutional neural network in optical remote sensing images[J]. International Journal of Distributed Sensor Networks, Vol. 15, No. 5, 2019.

[10] Shoulin Yin, Jing Bi. Medical Image Annotation Based on Deep Transfer Learning[J]. Journal of Applied Science and Engineering. Vol. 22, No. 2, pp. 385-390, 2019.

[11] Jing Yu, Hang Li, Shoulin Yin. Dynamic Gesture Recognition Based on Deep Learning in Human-to-Computer Interfaces [J]. Journal of Applied Science and Engineering, vol. 23, no. 1, pp.31-38, 2020.

[12] Kim Y . Convolutional Neural Networks for Sentence Classification[J]. 2014. arXiv:1408.5882.

[13] Hua C C, Qiu Y F, Wang Y B, et al. An augmented delays-dependent region partitioning approach for recurrent neural networks with multiple time-varying delays[J]. Neurocomputing, 2021, 423:248-254.

[14] Liu X, Zhou J, Qian H. Short-term wind power forecasting by stacked recurrent neural networks with parametric sine activation function[J]. Electric Power Systems Research, 2021, 192(4):107011.

[15] Zhang T, Zeng Y, Zhang Y, et al. Neuron type classification in rat brain based on integrative convolutional and tree-based recurrent neural networks[J]. Scientific Reports, 2021, 11(1).

[16] Francisco [illegible], Daniel R. [illegible] Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition[J]. Sensors, 2016, 16(1):115.

[17] Chen [illegible] Yu J, Zhao S, et al. User's Review Habits Enhanced Hierarchical Neural Network for Document-Level Sentiment Classification[J]. Neural Processing Letters, 2021(2).

[18] Shoulin Yin, Hang Li, Lin Teng, Man Jiang & Shahid Karim. An optimised multi-scale fusion method for airport detection in large-scale optical remote sensing images [J]. International Journal of Image and Data Fusion, vol. 11, no. 2, pp. 201-214, 2020. DOI: 10.1080/19479832.2020.1727573

[19] Xiaowei Wang, Shoulin Yin, Desheng Liu, Hang Li & Shahid Karim. Accurate playground localisation based on multi-feature extraction and cascade classifier in optical remote sensing images [J]. International Journal of Image and Data Fusion, vol. 11, no. 3. pp. 233-250, 2020. DOI: 10.1080/19479832.2020.1716862

[20] Jamshidi S, R Azmi, Sharghi M, et al. Hierarchical deep neural networks to detect driver drowsiness[J]. Multimedia Tools and Applications, 2021:1-14.

[21] Cao Y, Jia M, Ding P, et al. Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network[J]. Measurement, 2021, 178(5):109287.

[22] Ullah A, Muhammad K, Ding W, et al. Efficient Activity Recognition using Lightweight CNN and DS-GRU Network for Surveillance Applications[J]. Applied Soft Computing, 2021, 103(12).

[23] Londt T, Gao X, Xue B , et al. Evolving Character-level Convolutional Neural Networks for Text Classification[J]. 2020. arXiv:2012.02223

[24] J. W. Jang, Y. C. Kwon, H. Lim and O. Choi, "CNN-Based Denoising, Completion, and Prediction of Whole-Body

Human-Depth Images," in IEEE Access, vol. 7, pp. 175842-175856, 2019, doi: 10.1109/ACCESS.2019.2957862.

[25] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza and F. Pla, "Deep Pyramidal Residual Networks for Spectral–Spatial Hyperspectral Image Classification," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 2, pp. 740-754, Feb. 2019, doi: 10.1109/TGRS.2018.2860125.

[26] Shibata C, Uchiumi K, Mochihashi D. How LSTM Encodes Syntax: Exploring Context Vectors and Semi-Quantization on Natural Text[C]// Proceedings of the 28th International Conference on Computational Linguistics. 2020.

[27] X. Liang, L. Lin, X. Shen, J. Feng, S. Yan and E. P. Xing, "Interpretable Structure-Evolving LSTM," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2175-2184, doi: 10.1109/CVPR.2017.234.

[28] Hackel T, Usvyatsov M, Galliani S, et al. Inference, Learning and Attention Mechanisms that Exploit and Preserve Sparsity in Convolutional Networks[J]. International Journal of Computer Vision, 2020, 128(4).

[29] Xiaowei Wang, Shoulin Yin, Ke Sun, Hang Li, Jie Liu and Shahid Karim. GKFC-CNN: Modified Gaussian Kernel Fuzzy C-means and Convolutional Neural Network for Apple Segmentation and Recognition [J]. Journal of Applied Science and Engineering, vol. 23, no. 3, pp. 555-561, 2020.

[30] Hu Z, Zhang Z, Zhe S, et al. Salient object detection via sparse representation and multi-layer contour zooming[J]. Iet Computer Vision, 2017, 11(4):309-318.

[31] Ouyang Y, Zeng Y, Gao R, et al. Elective feature: The influence factor mining of students' graduation development based on hierarchical attention neural network model with graph[J]. Applied Intelligence, 2020(3).

[32] Shoulin Yin, Hang Li, Desheng Liu and Shahid Karim. Active Contour Modal based on Density-oriented BIRCH Clustering Method for Medical Image Segmentation [J]. Multimedia Tools and Applications. Vol. 79, pp. 31049-31068, 2020.

[33] S. Yin and H. Li. Hot Region Selection Based on Selective Search and Modified Fuzzy C-Means in Remote Sensing Images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 5862-5871, 2020, doi: 10.1109/JSTARS.2020.3025582.

[34] Xin Sun, Zheng Tang, Yongyan Zhao, Yingjie Zhang. Hierarchical Networks with Mixed Attention for Text Classification[J]. Journal of Chinese Information Processing, 35(2), 69-77, 2021.

[35] M. Rmayti, Y. Begriche, R. Khatoun, L. Khoukhi and D. Gaiti, "Denial of service (DoS) attacks detection in MANETs using Bayesian classifiers," 2014 IEEE 21st Symposium on Communications and Vehicular Technology in the Benelux (SCVT), 2014, pp. 7-12, doi: 10.1109/SCVT.2014.7046699.

[36] Chen Jin, Luo De-lin and Mu Fen-xiang, "An improved ID3 decision tree algorithm," 2009 4th International Conference on Computer Science & Education, 2009, pp. 127-130, doi: 10.1109/ICCSE.2009.5228509.

[37] Saishuai Zhao et al., "Extraction of mangrove in Hainan Dongzhai Harbor based on CART decision tree," 2014 22nd International Conference on Geoinformatics, 2014, pp. 1-6, doi: 10.1109/GEOINFORMATICS.2014.6950800.

[38] Jisi A and Shoulin Yin. A New Feature Fusion Network for Student Behavior Recognition in Education [J]. Journal of Applied Science and Engineering. vol. 24, no. 2, pp.133-140, 2021.

[39] Dan Zheng, Lei Meng, Shoulin Yin, Hang Li*. Enhanced Differential Privacy Protection Method Based on Adaptive Iterative Wiener Filtering in Discrete Time Series [J]. International Journal of Network Security. Vol. 23, No. 2, pp. 350-358, 2021.

[40] Desheng Liu, Linna Shan, Lei Wang, Shoulin Yin, et al. P3OI-MELSH: Privacy Protection Point of Interest Recommendation Algorithm Based on Multi-exploring Locality Sensitive Hashing[J]. Frontiers in Neurorobotics, 2021. doi: 10.3389/fnbot.2021.660304.

[41] Laghari, A.A., Wu, K., Laghari, R.A. et al. A Review and State of Art of Internet of Things (IoT). Arch Computat Methods Eng (2021). https://doi.org/10.1007/s11831-021-09622-6

[42] Laghari A A, Laghari M A. Quality of experience assessment of calling services in social network[J]. ICT Express, 2021, 7(2): 158-161. doi: 10.1016/j.icte.2021.04.011

[43] A. A. Laghari, H. He, A. Khan, N. Kumar and R. Kharel, "Quality of Experience Framework for Cloud Computing (QoC)," in IEEE Access, vol. 6, pp. 64876-64890, 2018, doi: 10.1109/ACCESS.2018.2865967.