Text Classification Based on Neural Network Fusion

Deageon Kim

Abstract: The goal of text classification is to identify the category to which the text belongs. Text categorization is widely used in email detection, sentiment analysis, topic marking and other fields. However, good text representation is the point to improve the capability of NLP tasks. Traditional text representation adopts bag-of-words model or vector space model, which loses the context information of the text and faces the problems of high latitude and high sparsity,. In recent years, with the increase of data and the improvement of computing performance, the use of deep learning technology to represent and classify texts has attracted great attention. Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and RNN with attention mechanism are used to represent the text, and then to classify the text and other NLP tasks, all of which have better performance than the traditional methods. In this paper, we design two sentence-level models based on the deep network and the details are as follows: (1) Text representation and classification model based on bidirectional RNN and CNN (BRCNN). BRCNN's input is the word vector corresponding to each word in the sentence; after using RNN to extract word order information in sentences, CNN is used to extract higher-level features of sentences. After convolution, the maximum pool operation is used to obtain sentence vectors. At last, softmax classifier is used for classification. RNN can capture the word order information in sentences, while CNN can extract useful features. Experiments on eight text classification tasks show that BRCNN model can get better text feature representation, and the classification accuracy rate is equal to or higher than that of the prior art. (2) Attention mechanism and CNN (ACNN) model uses the RNN with attention mechanism to obtain the context vector; Then CNN is used to extract more advanced feature information. The maximum pool operation is adopted to obtain a sentence vector; At last, the softmax classifier is used to classify the text. Experiments on eight text classification benchmark data sets show that ACNN improves the stability of model convergence, and can converge to an optimal or local optimal solution better than BRCNN.

Keywords: attention mechanism; deep learning; neural network; presentation; text classification

1 INTRODUCTION
1.2 Background and Significance

Today, people are more and more accustomed to frequently expressing their views and sharing their lives in the online world, involving a wide range of contents. In order to satisfy the users' expression needs, Tik Tok, Weibo, Tieba, etc. seized the opportunity, accumulated a mass of users, and produced a wealth of data with rich contents [1-3]. This kind of data grows and updates very fast and there are many ways to obtain it. The problem that people have to solve is no longer the acquisition of information, but the extraction of useful information. How to extract information quickly and effectively will become an important research topic. Compared with sound and image as carriers of information, text uses less network resources and is easy to transmit [4-6]. Effective understanding and analysis of the profound meaning of a text is called text representation. The traditional text representation method is to manually select features to label large-scale original text sets, which is tedious and difficult to achieve the desired results, and even more difficult to meet the information processing needs of users. How to make users get the required information accurately has become the focus of research. Text classification can not only truly reflect the text information by finding a suitable structured text representation method, but also have corresponding distinguishing ability for different texts [7-10]. Text classification automatically marks texts according to some classification systems or a standard, which is common in the fields of text subject, emails detection, public opinion analysis and so on, and has important practical significance in the efficient management and effective utilization of text information. The traditional text representation method has irreparable shortcomings, which makes it difficult to achieve the expected performance in

practice. Therefore, the research of text representation and text classification is a subject with great theoretical value and practical needs. Since 1990s, statistical machine learning has made great progress in the research of part-of-speech tagging, syntactic analysis, named entity recognition and other topics. Traditional text representation methods have gradually become the mainstream text representation methods. However, BOW's transformation of words into one-hot vectors ignores word order information and cannot reflect the internal relationship between words, which leads to semantic loss. TF-IDF suppresses the weight of high-frequency words in order to highlight low-frequency words, but exaggerates the importance of uncommon words. Common words are not equal to meaningless words. Text Rank's graph-based sorting algorithm extracts keywords to represent the text by voting, which is better for long texts than short texts, and the semantic loss of the text representation results is serious. LDA uses the topic distribution of the text and words in the topic to extract keywords to represent the text, which will also cause serious information loss, and has higher requirements for the topic dictionary. In order to solve the problems such as the lack of text information, specific keywords are often artificially selected to supplement the text information according to the specific situation. Although it has a certain effect, it makes the application scope of the model cramped and difficult to popularize, and it is even more difficult to face massive data. Common machine learning classification methods are: Logistic Regression (LR), which is prone to under-fitting or over-fitting, and the classification accuracy, is not ideal; Support Vector Machine (SVM)'s order is equal to the number of training samples. When there are many samples, it will be very difficult to calculate and store the matrix; KNN algorithm, the value of super parameter K has a great influence on the classification results. KNN has a high demand for data distribution, and it

wants the data to be clustered, so the classification effect is very poor when dealing with spiral data.

As for sentence-level text representation, with the increase of data and the improvement of computing performance, the use of deep learning techniques to represent and classify texts has attracted much attention. For example, NLP tasks such as using CNN [11-14], RNN [15-18, 26] and attention mechanism [19, 20] to represent documents and then classify texts have better performance than traditional methods. These neural network-based text representation and classification methods have been proved to be very effective. This is very important to deal with massive data in text form. It reduces the labor cost and improves ext classification accuracy and speed. However, compared with the achievements of deep neural network model in the area of computer vision, the application of deep neural network in the area of NLP is just beginning. In image classification, compared with the traditional machine learning method, the deep neural network model can reduce the error by more than 10%. In the competition of ImageNet Large Scale Recognition Association, the error of the deep neural network model has been reduced to 3.57%, which has exceeded the recognition ability of people. However, in natural language processing, such achievements can't be achieved at present. In recent years, experts and scholars in the field of natural language processing are also looking for ways to use deep neural network model to improve the performance of natural language processing tasks. Therefore, it is still a worthy research direction to use deep neural network for NLP.

1.3 Related Work
1.3.1 Research Status of Text Representation

In the long history of the development of human characters, characters are gradually different from sounds and images. Characters are no longer a simple signal, but a very abstract concept. To understand human words, we can't rely on logic alone, but also need a very strong knowledge base. The essence of text is a collection of a large number of words, which is accompanied by information such as spoken language, popular words, abbreviations, spelling mistakes and even emojis. It is extremely difficult for the existing text classification algorithms to directly identify this set, so it is necessary to convert it into a unified representation that can be easily recognized by the text classifier, that is, text representation. In 2005, Zhou proposed a graph-based text representation method. The text is converted into text features according to the established rules, and these features are synthesized to define the similarity used to calculate the measurement chart. In this method, the word order relationship in the text is taken into account, but many manually set parameters are included in the configuration process. In 1998, Salton put forward the Vector Space Model (VSM), which uses vectors to represent text features. Related research issues mainly focus on feature selection and weight calculation. Set a threshold value for each feature, remove the features whose support is less than the threshold value, and regard the rest features as valid features. Common feature

selection methods are: chi-square statistics, text frequency, mutual information, expected cross entropy, etc. Usually, the corresponding weights are calculated based on the frequency of the selected features. VSM ignores the relationship between words and text contexts, which leads to the loss of text information. In recent years, although the Bag of Word (BOW) model has been widely used in text representation, BOW defaults that each word is independent, ignoring the relationship between words and the information between text contexts. Moreover, the expression vector of words is generally in the form of one-hot coding vector, and the vector dimension is the size of the bag of words, which makes the result of text feature representation face the problem of dimension disaster and high sparseness. Word embedding model gradually replaces word bag model, which is used for vector representation of words. The essence of word vector is the first full connection layer parameter of deep neural network, in which unique hot coding is used as input. The closer the word meaning is, the higher the similarity of the corresponding word vectors in feature space, and vice versa. Although the word embedding model solves the problem of high sparseness and reduces the vector dimension compared with the bag of words model, the dimension is still very high, which makes it difficult to show its advantages in traditional classification algorithms. In 2002, Lai proposed the RCNN model, using the advantages of RNN and CNN. Because RNN itself is a sequence model, for the text feature sequences represented by word vectors, RNN processing can extract the context information of the text well, and CNN processing can obtain a number of local information and greatly reduce the amount of calculation, so RCNN has achieved good results in text classification; In 2014, Shen embedded the text representation based on word vectors, and combined with CNN to mine the high-level semantics of the text; In 2014, Santos proposed DCNN model for sentiment classification based on word vectors of characters, words and sentences. The model achieved good results, but it was difficult to build a depth model when the text length was not fixed. To sum up, appropriate text representation and appropriate depth model are the basis of text classification.

1.3.2 Research Status of Text Classification

Text classification is a process of labeling each sample with appropriate category according to relevant features. Text classification methods are mainly based on pattern system or machine learning [27]. The classification method based on pattern system uses knowledge engineering technology and professional assistance to design appropriate classification rules for each category. If the text conforms to the corresponding rules, it is considered that the text belongs to that category. The text features of knowledge engineering refer to the related attributes of the established rules in the text. Due to the role of artificial judgment, a good accuracy rate is obtained, but the defects of the pattern system are obvious. The classification rules determine the quality of the classification results, and the formulation of rules requires a lot of research and demonstration by experts. However, these rules cannot be used across fields, or even transferred to

different tasks in the same field, which leads to the basic lack of popularization of knowledge engineering and restricts knowledge. Classification models based on statistical machine learning, including Naive Bayes (NB), KNN, SVM, neural network, decision tree, etc. [21, 22]. The classification performance of NB, SVM, decision tree and KNN is poor, so some people improve it. Zhou Zhihua et al. put forward the selective integration theory in 2002, and proved the superiority of integrated learning system, and achieved good results in text classification. However, the above models are all shallow machine learning methods, which can deal with simple classification problems. Its generalization ability is often not strong, and the phenomenon of over-fitting or under-fitting often occurs in the face of high-dimensional feature data. Therefore, it is a feasible direction to explore the application of deep model in classification.

1.3.3 Research Status of Deep Neural Networks

DL (Deep learning) combines low-level features with nonlinear transformation to form more abstract high-level features, so that the model can better learn the distribution law of data [23]. In 1986, Rumelhart and Hinton proposed the back propagation algorithm, which made the neural network change from a simple model to a complex model, making great contributions to the development of deep learning. In 1998, Lecun et al. used convolution neural network to reduce parameters and calculations and improve the performance of model training by using local perception and weight sharing. In 2000, Hinton put forward the learning algorithm of contrast divergence, and in 2006, he put forward the Restricted Boltzman Machine (RBM). Through layer-by-layer training, the problem of deep learning model optimization is solved, which makes DL develop rapidly. In 2007, Alex Graves et al., based on Long Short-Term Memory (LSTM), recognized handwritten characters, and Sutskever et al., in 2014, proposed to build a machine translation framework by using two multilayer LSTM network structures. In 2008, Vincent proposed Auto-Encoder (AE), which is a neural network to reconstruct the input information, hoping that the input is equal to the output. Its basic structure consists of input layer, hidden layer and output layer. The input data is the original data. After the change of the hidden layer, the output results are as consistent as possible with the input data. Then the parameter matrix of the hidden layer can be used as the characteristics of the original data to achieve better results. Some researchers have made some improvements to AE. The denoising self-coding improves the anti-interference ability of data, and the variational self-coding changes the sample distribution to generate new samples. In 2011, Socher and others proposed a Recursive Neural Network to predict the tree structure. In 2012, when DL was used in Imag Net tasks, the error rate dropped from 26% to 15%. DL developed from the field of speech to the field of image recognition in academic circles. In 2013, Mikolov proposed sequence-based deep neural network RNN [24, 25]. The model ingeniously adds self-connection and interconnection in the hidden layer, and has certain memory ability. In 2015, Hinton et al. said on Nature:

"In the next few years, DL will have a huge impact on the field of natural language understanding". In 2017, Young et al. compared DL models in various NLP fields and analyzed the possible trends in the future.

1.3.4 Research Status of Deep Learning in NLP

NLP is another important application field of deep learning. In the field of NLP, the use of neural network model for word embedding, RNN and CNN for text classification and translation tasks have made great progress. In 2000, Xu first put forward the idea of using neural network to train language model. A language modeling method using three-layer neural network is introduced in detail in the literature. Literature puts forward the idea of hierarchy to replace the matrix multiplication from hidden layer to output layer in the method of literature, which reduces the amount of calculation when the effect is equal to that of literature. Colobert and Weston introduced their word vector calculation method in, and systematically introduced their work in. Huang et al. improved the model proposed by Colobert et al. in, using full-text information to assist local information. Another innovation of Huang et al. is learning multiple word vectors to represent polysemous words. Mikolov proposed CB0W and Skip-gram models, and in proposed negative sampling and hierarchical softmax to improve the calculation speed of the models. Rong analyzed the learning process of the parameters of Mikolov's word vector model in detail in. Jeffrey et al. put forward a method of word vector representation by decomposing the global "word-word" co-occurrence matrix in Glove model. These methods are all based on the distribution representation of distribution hypothesis, which can map words into low-dimensional vector space without losing the correlation between words. They are the most advanced word embedding methods at present. The method of text representation and classification at sentence level or document level is also a research hotspot. Hill et al. proposed to learn the distributed representation of sentences from unlabeled data. Conneau et al. proposed using supervised learning to learn general sentence representation from natural language reasoning data. And lemikolov in proposed learning paragraph vectors for classification. Illinois-LH is a distribution table system for learning semantics. Kim proposed a word-level CNN text classification model. Sochei et al. proposed using deep RNN model for text classification. Dai and Lip came up with using semi-supervised learning to improve sequence learning. Kalchbrenner et al. used CNN to model sentences. Skipthouht proposed by Kiros et al. and F-Dropout proposed by Wang and Manning have achieved good performance. Machine translation is one of the best tasks combining deep learning with natural language processing, and many machine translation models based on deep neural networks have achieved good performance. Cho et al. used RNN encoder-decoder to learn the phrase representation for statistical machine translation. Cho et al. in introduced the properties of the encoder-decoder method of neural machine translation. Devlin et al. applied neural network language model to the decoding process of machine translation

framework. Sutskever et al. used the SEQ 2 SEQ model of a single neural network. On the basis of sequence-to-sequence machine translation model, Bahau et al. added attention mechanism to automatically learn the aligned words of the target words in the source language. Dong et al. used the multi-task learning method to jointly learn the translation model from one source language to multiple target languages. Meng et al. used neural turing machine to build a deep memory neural network. Kalchbrenner et al. introduced a neural machine translation model with linear time complexity. These translation models based on deep neural networks have achieved better performance than traditional translation models. To sum up, in general, the achievements of DL in NLP are far from those in image and speech recognition, and there is still much room for research.

2 RELATED WORK
2.1 Text Classification Method

Text classification refers to the process of classifying a large number of texts into one or more categories. Text classification is not much different from general classification problems. It is based on the characteristics of the samples to be classified to judge and choose the best classification result. The earlier text classification method is the matching method, which focuses on whether there are the same words or words with the same meaning in the text [28]. This factor is used to analyze the category of the text. This method is too simple and narrow to get satisfactory classification results. At present, the mainstream text classification method is machine learning method based on statistics, which takes the data set of known categories as the training set. According to the effective training classifier in the training set, the text of unknown category is classified and predicted. Compared with knowledge engineering technology, statistical learning method has been used as a common method for classification problems, because there are a large number of technologies with detailed theoretical basis, and there are many subjective factors of knowledge engineering experts. Traditional classification methods include KNN, decision tree, ensemble learning, etc.

2.2 Overview of Deep Neural Networks

Deep learning is a method of learning data distribution, which is mainly used in interpreting and analyzing image, sound and text data. In recent years, deep learning has made great achievements, and then it has been widely studied and applied. Here are some common deep learning models.

2.2.1 RNN

Recurrent Neural Networks consists of 3 layers, input layer (x), hidden layer (s) and output layer (o). As shown in Fig. 1, self-connection and interconnection are added to the hidden layer of its model, which enables RNN to acquire short-term memory properties and be widely used to process sequence features.
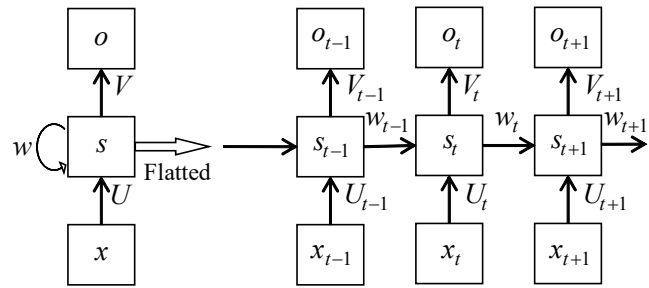
Figure 1 Basic structure of rnn network

Each node in Fig. 1 represents a unit. RN is a sequence model, x_{t-1}, x_t, x_{t+1} is an input sequence, o_{t-1}, o_t, o_{t+1} is an output sequence and s_{t-1}, s_t, s_{t+1} is a memory sequence of the hidden layer.

$$S_t = f\left(U * X_t + U * X_{t-1}\right) \tag{1}$$
$$O_t = soft\max\left(VS_t\right) \tag{2}$$

For the node at time t of RNN:
(1) S_t is in a hidden state, capturing the information on the node at the previous moment.
(2) o_t is obtained from the memory of the current time node and all previous time nodes.
(3) However, s_t can't capture the information of all nodes after T time.
(4) Each memory cell shares a set of parameters (U, V, w), which greatly reduces the amount of calculation.
(5) In many cases, o_t is not output, only the result of the last moment of the sequence is output.

However, there is not only one common network structure of RNN. Several common RNN structures are shown in Fig. 2.
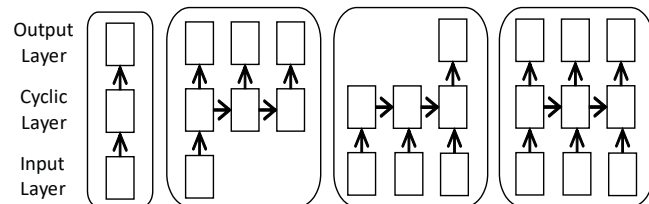
Figure 2 Common structure of cyclic neural network

In addition, RNN can be divided into two types: static and dynamic, depending on whether the number of cells needs to be set in advance. The static circular neural network requires that the number of hidden units of the preset circular neural network should not be changed, and the length of the input text sequence should be determined strictly according to the number of hidden units, that is to say, the length of the input text sequence must be consistent. Dynamic neural network doesn't need to set the number of hidden units of cyclic neural network in advance, but only automatically adjusts hidden units' amount according to the length of input sequence, that is to say, it doesn't need the same text length. However, dynamic RNN is not very effective in dealing with the problems of long text and text coding and decoding.

2.2.2 CNN

CNN has achieved great success in computer vision, and it has been deeply studied and widely used. From Lenet, Alex Net, VGG Net, Inception (Google Net) and later Rest Net, the network has become deeper and deeper, but its performance has become more and more prominent. In recent years, CNN has become popular in the field of text processing. The essence of convolution is an integral transformation. Generally, multiple convolution kernels are often used in convolution operation, which makes the number of channels of the obtained feature matrix more than that of the original feature matrix. After finishing, the dimension of the feature matrix will become larger and larger, which not only does not simplify the problem, but also requires more parameters and more memory in the subsequent calculation. Therefore, pooling operation is widely used to solve this problem. Pooling is used in the non-overlapping areas of the matrix, which is equivalent to an abstract process, filtering out unnecessary information, and generally, there are three kinds of pool operations:
(1) Maximum pool: the maximum value in the pool window is regarded as the value of this area;
(2) Average pooling: the average value in the pool window is regarded as the value of this area.
(3) Random pool: randomly select the values in the pool window.

Among them, the largest pool operation is the most widely used in text processing.

3 EXPERIMENTAL PROCESS
3.1 Experimental Principle

Usually, the CNN is composed of multiple convolutional layers, and each convolutional layer usually performs the following operations: Pool the output result of the activation function (usually using the maximum pool operation) to get the most significant features,

These steps constitute a common convolution layer, and sometimes local response normalization (local re-spone normalization (LRN) operation. The superposition of multiple convolution layers can obtain higher-level features. Brcnn model firstly combines loop operation, convolution operation and pool operation to extract the features of the text and express the text as feature vectors; then, the full connection layer is used to transform the feature space, and then the Softmax classifier is used for classification. The model structure diagram is shown in Fig. 3.

The input of the text classification part is the sentence vector M obtained from the representation part, which is mainly used to predict the category of the sentence.

(1) Full connection layer feature space conversion

Here, a fully connected layer is used to transform the feature space.

$$O_f = relu\left(W_f^T m + b_f\right) \tag{3}$$

The input of the whole connection layer is the sentence vector learned at the end of the convolution layer $m \in R^M$, and the output is a transformation with the same dimension as the input $O_f \in R^M$. $W_f \in R^{M \times M}$ and $b_f \in R^M$ are available parameters for learning. The activation function used here is *relu*. To prevent over-fitting, drop operation is added.

(2) Output layer classification result prediction

There is a full connection mapping from the full connection layer to the output layer:

$$y = W_o^T O_f + b_o \tag{4}$$

The input of this layer is the output of the full connection layer, and the output is the predicted value belonging to each category $y \in R^C$, where C is the number of target categories $W_o \in R^{M \times C}$, $b_o \in R^C$ is the weight available for learning. To have an intuitive understanding, use softmax to normalize y to get the predicted probability of each category $p \in R^C$, where the formula of the ith component p_i in p is as follows:

$$P_i = \frac{y_i}{\sum_{c=1}^{C} y_c}. \tag{5}$$

3.2 Data Set

In this experiment, eight benchmark data sets are used, as shown in Tab. 1: C corresponds to the number of target categories in the data set, L corresponds to the average length of sentences in the data set, N corresponds to the number of samples in the data set, V is the vocabulary size of the data set, and Test is the sample number of the test set, and CV means cross-validation with 10% discount.

Table 1 A summary of the datasets

Data Set	C	L	N	V	Test
MR	3	19	10,660	18,765	CV
Subj	2	24	10,000	21,323	CV
SST	6	17	11,860	17,836	2,210
SST2	3	19	9,600	16,185	1,821
IMDB	3	231	50,000	39,200	25,000
TREC	7	11	5,952	9,125	500
CR	2	19	3,741	5,340	CV
MPQA	2	3	10,504	6,246	CV

The specific description of each data set is as follows:

MR: MR is movie review data marked by Pang et al. for emotion classification.

Set, which was first used in reference. Each sentence in the data set corresponds to a comment. This data set is divided into positive/negative reviews, and it is a two-category task data set, in which there are 5331 positive examples and 5331 negative examples, with a total of 10662 reviews. The average length of sentences in this data set is 20, and the vocabulary size is 18,765 words. In this experiment, 10-fold cross-validation is used.

Subj: sub (objectivity) is a sentiment analysis data set containing 5,000 subjective and 5,000 objective sentences, which was used for the first time by Pang et al. in literature. This is also a data set with two target classes (subjective and objective :), that is, a binary data set. Similarly, each sentence corresponds to a sample. The average sentence length is 23, and the number of vocabulary words is 21,323. In the experiment of this paper, the 10-fold cross-validation is also used.

SST: SST (Stanfod Sentimental Treebank) is marked and published by Socher et al., which is an extension of MR. This data set includes a total of 11,855 movie reviews, which are labeled as five categories (very positive, positive, neutral, negative and very negative), that is, it is a five-category data set. This data set provides the segmented training set (8544), calibration set (1101) and test set (2210).

SST: In SST2, remove neutral comments in SST, and combine very positive and positive into positive, and very negative and negative into negative. The final SST2 contains a total of 9163 samples, including 7792 samples in the training set and 1821 samples in the testing set. Similarly, this is a binary classification task with two target classes.

MDB: IMDB data set is a two-class sentiment analysis data set, including 50,000 samples, 25,000 samples in training set and 25,000 samples in testing set, which was marked and published by Maas et al. This is a binary data set, which contains more data than the previous benchmark data set. Additional unlabeled data can also be used.

TREC: TREC is the problem classification task data set T97 marked by Li et al. This data set is divided into training set and testing set and the training set is randomly divided into training sets of 1000, 2000, 3000, 4000 and 5500 samples. This paper uses the training set with 5500 training samples.

CR: CR is customer reviews M1 of 14 products obtained from Amazon annotated by Humin Qing et al. Its task is to classify each customer's review into positive and negative categories.

3.3 Comparison and Analysis

This section introduces the comparison and analysis of experimental results in detail. Including BRCNN and ACNN and existing modules

Performance comparison and analysis of BRCNN, ACNN and their variants, multilayer.

Comparison and analysis of performance of BRCNN and ACNN, comparison and analysis of dropout strength of BRCNN circulating layer. Comparison and analysis of accuracy, recall and F1 value, P-R curve analysis, confusion matrix analysis, ROC curve and AUC analysis, etc. Tab. 2 introduces BRCNN and ACNN, their variants and existing models in eight benchmark data sets.

Tab. 2 shows the comparison of the accuracy of BRCNN and ACNN and their variants with the existing models on MR, SUBJ, TREC, CR and MPQA data sets. BRCNN and ACNN based on RNN have almost the same accuracy rate as the existing methods, while BRCNN and ACNN based on LSTM and GRU have the same or higher accuracy rate as the

existing models. Particularly, on the TREC data set, the BRCNN based on LSTM designed in this paper reduces the error rate by 23.4%; GRU-based BRCNN reduces the error by 26.7%; ACNN based on RNN reduces the error rate by 6.3%; ACNN based on LSTM reduces the error by 34.4%; ACNN based on GRU also reduced the error rate by 26.7%. For CR data sets, BRCNN based on LSTM reduces the error rate by 7.3%; ACNN based on LSTM also reduces the error by 8.8%. ACNN based on LSTM reduces the error by 1.8% on MR data sets, while BRCNN based on LSTM also achieves the second highest performance. As for Subj data set and MPQA data set, BRCNN and ACNN also got the second highest accuracy.

What is listed in Tab. 3 is the comparison of the correct rates on SST data set and SST2 data set. Br CNN and ACNN also got the second highest and the third highest accuracy respectively. SST data sets are divided into five categories. On the other hand, SST2 data set is a binary classification task based on SST data set, which removes the neutrality and combines the positive and negative data sets. Therefore, the classification features are more obvious, so the classification results are better than SST data set.

Table 2 Results of our models and the state-of-the-art models on MR> Subj TREC CR and MPQA datasets

Model	MR	Subj	TREC	CR	MPQA
BRCNN(RNN)	81.6	93.1	92.4	80.0	90.3
BRCNN(LSTM)	83.1	94.2	95.1	87.3	90.7
BRCNN(GRU)	82.6	94.1	95.3	85.6	90.8
ACNN(RNN)	81.6	93.3	94.0	86.1	90.3
ACNN(LSTM)	83.4	95.2	95.8	87.5	90.5
ACNN(GRU)	82.7	94.1	95.3	86.9	90.8
FastSent	70.8	89.0	76.8	78.4	80.6
FastSent+AE	71.8	88.8	80.4	76.7	81.5
SkipThought	76.5	93.6	92.2	80.1	87.1
SkipThought-LN	79.4	93.7	88.4	83.1	89.3
ParagraphVec(DBOW)	60.2	76.3	59.4	66.9	70.7
SDAE+embs.	74.6	90.8	78.4	78.0	86.9
Unigram-TFIDF	74.7	90.3	85.0	79.2	82.4
CNN-static	81.0	93.0	92.8	84.7	89.6
CNN-non-static	81.5	93.4	93.6	84.3	89.5
CNN-multichannel	81.1	93.2	92.2	85.0	89.4
Paragraph-Vec	75.8	90.5	91.8	78.1	74.2
MNB	79.0	93.6	-	80.0	86.3
word2vecBOW	77.7	90.9	83.6	79.8	88.3
fastTextBOW	76.5	91.6	81.8	78.9	87.4
GloVeBOW	78.7	91.6	83.6	78.5	87.6
GloVePositionalEncoding	78.3	91.1	83.3	77.4	87.1
BiLSTM-Max(untrained)	77.5	89.6	85.8	81.3	88.7
CaptionRep(bow)	60.9	77.4	72.2	69.3	70.8
DictRepBOW+embs.	76.7	90.7	81.0	78.7	87.2
NMTEn-to-Fr	64.7	84.9	82.8	70.1	81.5
BiLSTM-Max(onSNLI)	79.9	92.1	88.7	84.6	88.8
BiLSTM-Max(onA11NLI)	81.1	92.4	88.2	86.3	90.2
NaiveBayes-SVM	79.4	93.2	87.4	81.7	86.3
RNN	77.2	82.3	90.2	82.3	90.1
BRNN	82.3	94.2	91.0	82.6	90.3
AdaSent	83.1	95.5	92.4	86.3	93.3
G-Dropout	79.0	93.4	93.4	82.1	86.1
F-Dropout	79.1	93.6	89.7	81.8	86.3

What is shown in Tab. 4 is the comparison of the correct rates on IMDB data sets. The accuracy of BR-CNN and ACNN on IMDB data sets is 91.1%, which is higher than

most existing models. Compared with paragraph-Vec, SA-LSTM and SEQ 2-bowen-CNN, the accuracy performance of BRCNN and ACNN is not good enough. As can be seen from Tab. 4, the average sentence length of IMDB data set is 231. And in IMDB data set, a sample corresponds to multiple sentences. BRCNN and ACNN are more effective for single-sentence short texts, while paragraph and other methods are designed to handle long texts, so they have better performance on IMDB data sets than BRCNN and ACNN. BRCNN and ACNN can also get higher accuracy on data sets that they are not good at, so BRCNN and ACNN can still extract sentence features very well. To sum up, the models BRCNN and ACNN designed in this paper can well extract the feature information of sequence data, convert sentences into corresponding vector representations, and get better sentence representations, so as to achieve higher or equivalent accuracy than the existing models in the classification task. This shows the feasibility of BRCNN and ACNN.

Table 3 Accuracy of models on the SST and SST2 datasets

Model	SST	SST2
BRCNN(RNN)	463	86.0
BRCNN(LSTM)	49.0	87.8
BRCNN(GRU)	47.6	87.4
ACNN(RNN)	46.6	86.3
ACNN(LSTM)	49.1	87.8
ACNN(GRU)	48.9	87.5
SkipThought	47.9	82.0
SkipThought-LN	48.3	82.9
CNN-static	45.5	86.8
CNN-non-static	48.0	87.2
CNN-multichannel	47.4	88.1
RCNN	47.2	87.6
RNTN	45.7	85.4
DCNN	48.5	86.8
Paragraph-Vec	48.7	87.8
word2vecBOW	47.6	79.7
fastTextBOW	46.7	78.8
GloVeBOW	45.5	79.8
GloVePositionalEncoding	43.6	80.6
BiLSTM-Max(untrained)	44.9	80.7
BiLSTM-Max(onSNLI)	42.5	83.3
BiLSTM-Max(onA11NLI)	43.3	84.6
NaiveBayes-SVM	45.6	83.1

Table 4 Accuracy of models on the IMDB dataset

Models	Accuracy
BRCNN(RNN)	88.0
BRCNN(LSTM)	91.1
BRCNN(GRU)	91.0
ACNN(RNN)	88.4
ACNN(LSTM)	91.5
ACNN(GRU)	93.1
MNB	86.9
G-Dropout	91.2
F-Dropout	93.1
NBSVM	91.2

Tab. 5 is a comparison of the performance of multi-layer BRCNN and multi-layer ACNN on TREC data sets. With the increase of layers, the accuracy of BRCNN increases, but that of ACNN hardly increases. However, ACNN on the first layer has achieved the same accuracy as that of multi-layer

BRCNN. This further shows that ACNN can extract the features of the text better and get a better sentence feature representation (that is, get a better sentence vector), so it can get a better classification result in the classification task.

Table 5 Accuracy of ACNN and BRCNN models on the TREC dataset

Models	Accuracy
BRCNN(1 Bi-LSTM layer)	95.5
BRCNN(2 Bi-LSTM layers)	95.7
BRCNN(3 Bi-LSTM layers)	95.7
ACNN(1 Bi-attention layer)	95.7
ACNN(2 Bi-attention layers)	95.7
ACNN(3 Bi-attention layers)	96.3

4 CONCLUSION

Text classification has always been a hot topic in NLP field, and how to express text into digital features is the key of text processing. The main work of this paper is also devoted to the research of text representation and classification methods. In this paper, two text representation and classification models based on deep neural networks are designed: BRCNN and ACNN. The feasibility of BRCNN and ACNN is verified by experiments. This paper mainly designs a text representation and classification model based on Bidirectional RNN and CNN (BRCNN). BRCNN extracts word order information by using RNN, then uses CNN to extract higher-level features, and then uses maximum pooling operation to get sentence vectors, and embeds sentence features in vector space. Then use softmax classifier to classify. Finally, through experiments on eight benchmark data sets and comparative analysis with existing models, it can be found that BRCNN and ACNN can represent and classify texts well. It can also get good performance for data sets with unbalanced data. Even for long text, you can get good performance. In addition, by analyzing BRCNN, ACNN and their variants, it is concluded that the model based on LSTM and GRU can get better text representation than the model based on RNN, so the classification accuracy is higher. The convergence of LSTM and GRU-based models requires much less iteration rounds than that of RNN-based models. Comparing the accuracy of BRCNN and ACNN models, it is found that there is little difference between them. However, for BRCNN with multi-layer circulation layer and ACNN with multi-layer attention mechanism layer, the latter can stably converge to an optimal or local optimal solution.

5 REFERENCES

[1] Kim, J. N. (2019). The mediating effect of fear of missing out between interpersonal relationship orientation and SNS addiction tendency in college students. *Asia-pacific Journal of Convergent Research Interchange, SoCoRI, 5*(3), 1-9.

https://doi.org/10.21742/apjcri.2019.09.01

[2] Shin, B. & Kim, J. (2020). Chain-based routing model for improving the network lifetime in WSNs. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 6*(11), 173-181. https://doi.org/10.47116/apjcri.2020.11.15

[3] Kim, E. H. & Yilong, C. (2022). The effect of YanBang dishes selection attributes and sns information characteristics in china on brand loyalty: Focusing on the mediating effect of consumer satisfaction. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 8*(1), 37-51 https://doi.org/10.47116/apjcri.2022.01.04

[4] Wang, X. (2020). The construction of employment and entrepreneurship service system for university students based on big data. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 6*(10), 203-213. https://doi.org/10.47116/apjcri.2020.10.16

[5] Hong, Y., Shin, S., Kim, N., Kim, J., & Kim, H. (2022). A study on the perception based on big data analysis related to Osiria tourism complex in Busan. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 8*(3), 59-68 https://doi.org/10.47116/apjcri.2022.03.06

[6] Yoo, H. (2022). Analysis of research status of early childhood physical education using big data. Asia-pacific Journal of Convergent Research Interchange, FuCoS, 8(5), 97-106 https://doi.org/10.47116/apjcri.2022.05.09

[7] Kim, S. Y. (2018). Urban image formation through evaluation of public design guideline of bus stop in Jeju. Asia-pacific Journal of Convergent Research Interchange, SoCoRI, 4(2), 21-30. https://doi.org/10.14257/apjcri.2018.06.03

[8] Asish Vardhan, K. (2019). Some studies on digital image segmentation techniques. *Asia-pacific Journal of Convergent Research Interchange, SoCoRI, 5*(1), 77-89. https://doi.org/10.21742/apjcri.2019.03.08

[9] Trikha, V., Fiaidhi, J., & Mohammed, S. (2020). Identifying EEG binary limb motor imagery movements using thick data analytics. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 6*(9), 169-189 https://doi.org/10.47116/apjcri.2020.09.15

[10] Lee, J. Y. & Durbha, M. (2016). Customary broadcast encryption with advanced encryption and short ciphertexts. *Asia-pacific Journal of Convergent Research Interchange, SoCoRI, 2*(2), 27-33. https://doi.org/10.21742/APJCRI.2016.06.04

[11] Liu, T. & Yang, D. (2021). A three-branch 3D convolutional neural network for EEG-based different hand movement stages classification. *Scientific Reports, 11*(1), 10758. https://doi.org/10.1038/s41598-021-89414-x

[12] Li, G., Liu, F., & Wang, Y. (2021). A convolutional neural network (CNN) based approach for the recognition and evaluation of classroom teaching behaviour. *Scientific Programming, 2021*, Article ID 6336773. https://doi.org/10.1155/2021/6336773

[13] Li, T., Xu, Y., & Luo, J. (2021). A method of amino acid terahertz spectrum recognition based on the convolutional neural network and bidirectional gated recurrent network model. *Scientific Programming, 2021*, Article ID 2097257. https://doi.org/10.1155/2021/2097257

[14] Sang, X., Zhou, R. G., & Li, Y. (2021). One-dimensional deep convolutional neural network for mineral classification from Raman spectroscopy. *Neural Processing Letters, 54*(1), 677-690. https://doi.org/10.1007/s11063-021-10652-1

[15] Lee, S. H., Goau, H., & Bonnet, P. (2020). Attention-based recurrent neural network for plant disease classification. *Frontiers in Plant Science*, 11, Article 601250. https://doi.org/10.3389/fpls.2020.601250

[16] Ren, X., Gu, H., & Wei, W. (2021). Tree-RNN: Tree structural recurrent neural network for network traffic classification. *Expert Systems with Applications, 167*(1), 114363. https://doi.org/10.1016/j.eswa.2020.114363

[17] Jing, Z., Liu, J., & Ibrahim, M. S. (2020). Lifetime prediction of ultraviolet light-emitting diodes using a long short-term memory recurrent neural network. *IEEE Electron Device Letters, 41*(12). https://doi.org/10.1109/LED.2020.3034567

[18] Liu, Y., Al-Shehari, H., & Zhang H. (2022). Attention mechanism enhancement algorithm based on cycle consistent generative adversarial networks for single image dehazing. *Journal of Visual Communication and Image Representation, 83*, 103434. https://doi.org/10.1016/j.jvcir.2021.103434

[19] Zhang, H., Peng, G., & Wu, Z. (2022). MAM: A multipath attention mechanism for image recognition. *IET Image Processing*, 16. https://doi.org/10.1049/ipr2.12370

[20] Xu, X., Gao, T., & Wang, Y. (2022). Event temporal relation extraction with attention mechanism and graph neural network. *Tsinghua Science and Technology, 27*(1), 79-90. https://doi.org/10.26599/TST.2020.9010063

[21] Dehghan-Dehnavi, S., Fotuhi-Firuzabad, M., & Moeini-Aghtaie, M. (2020). Decision-making tree analysis for industrial load classification in demand response programs. *IEEE Transactions on Industry Applications*, (99), 1. https://doi.org/10.1109/TIA.2020.3032932

[22] Amgad, M., Atteya, L. A., Hussein, H., Hosny Mohammed, K., Hafiz, E., Elsebaie, M. A. T., Mobadersany, P., Manthey, D., Gutman, D. A., Elfandy, H., & Cooper, L. A. D. (2022). Explainable nucleus classification using Decision Tree Approximation of Learned Embeddings. Bioinformatics, 38(2), 513-519. https://doi.org/10.1093/bioinformatics/btab670

[23] Baek, M. & Baker, D. (2022). Deep learning and protein structure modelling. *Nature Methods, 19*(1), 13-14. https://doi.org/10.1038/s41592-021-01360-8

[24] Zhang, P., Xue, J., & Lan, C. (2020). EleAtt-RNN: Adding attentiveness to neurons in recurrent neural networks. *IEEE Transactions on Image Processing*, 29. https://doi.org/10.1109/TIP.2019.2937724

[25] Tian, Y., Yao, H., & Cai, M. (2021). Improving RNN transducer modeling for small-footprint keyword spotting. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021)*. https://doi.org/10.1109/ICASSP39728.2021.9414339

[26] Barak, F. & Kaplan, K. (2021). The study of handwriting recognition algorithms based on neural networks. *International Journal of Hybrid Innovation Technologies*, 1(2), 63-74. https://doi.org/10.21742/ijhit.2653-309X.2021.1.2.04

[27] Yang, L., Liguo, H., & Xuesen, C. (2015). A Small Sample Text Classification Algorithm based on Kernel Density Estimation, *International Journal of Signal Processing, Image Processing and Pattern Recognition, NADIA, 8*(6), 235-244. https://doi.org/10.14257/ijsip.2015.8.6.25

[28] Zhong, L. (2017). Semi-supervised Text Classification Using SVM with Exponential Kernel. *International Journal of Database Theory and Application, NADIA, 10*(1), 79-88. https://doi.org/10.14257/ijdta.2017.10.1.08

Author's contacts:

Deageon Kim
Architectural Engineering, Dongseo University,
47 Jurye-ro, Sasang-gu, Busan, 47011, Republic of Korea
gun43@hanmail.net