The Study on the Text Classification for Financial News Based on Partial Information

WENJIE ZHAO [ID][1], **GAOYU ZHANG** [ID][2], **GEORGE YUAN**[3,4], **JUN LIU**[5], **HONGTAO SHAN**[1], **AND SHUYI ZHANG** [ID][6]

[1]School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China
[2]School of Information Management, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China
[3]Business School, Chengdu University, Chengdu 610106, China
[4]School of Finbtech, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China
[5]Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China
[6]Lixin Research Institute, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China

Corresponding authors: Gaoyu Zhang (20089840@lixin.edu.cn) and Jun Liu (junliu@lixin.edu.cn)

ABSTRACT The goal of this paper is to conduct the study on the text classification for financial news based on partial information. By a fact that an indispensable step for the efficient use of topic information embedded in financial news is the text classification, a new neural network called "All Dataset based on CharCNN (Character Convolutional Neural Networks) and GRU (Gated Recurrent Unit)" (in short, AD-CharCGNN) which extracts a part of the financial article and incorporates both time domain and spatial domain to classify financial texts is proposed. In the study of this paper, we first build a character level vocabulary by reading all characters of the financial dataset, part of each financial text which will be classified is mapped to a high-dimensional spatial vector based on the vocabulary. Then, the vectors are convoluted in the spatial domain to get the text local features, and next, the features are processed by the gated recurrent units to get the features contained time information. Finally, the features which contain spatial and time information will be classified through softmax function to get the text classification results. Our results on the experiments confirm that the network proposed in this paper works effectively with the accuracy of 96.45%, and it seems that the text classification algorithm with the feature by taking only partial text part is more suitable for the application of the practice. Meanwhile, for the input with character level vector, the network is not only suitable for Chinese but also for other languages.

INDEX TERMS Financial news, natural language processing (NLP), text processing.

I. INTRODUCTION

We know that there is an interplay between financially related news and the financial market [1], and thus it has become an indispensable step to classify the massive internet financial news.

It is well known that the classification of financial news is the most fundamental step to help financial individuals or financial institutions to make decisions [2]. Especially for professional financial experts, after classification, detailed and effective financial texts can master the current advanced research and possible future directions, and comprehensively understand the financial information in network. The goal of this paper is to study and propose a

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano [ID].

new neural network called "All Dataset based on Char-CNN and GRU (AD-CharCGNN)" which extracts a part of the financial text and incorporates both time domain and spatial domain to classify financial news. In this paper, with 65,000 Chinese financial news crawled as a data set, the classification is accomplished with an accuracy of 96.45%.

In studies of the traditional classification algorithms, Kalra and Prasad applied the classifier Naive Bayes (NB) to categorize the financial news text. They proposed a daily prediction model and used the historical data and news articles to predict the stock market movements [3]. By combining Naive Bayes with the decision tree C4.5 algorithm, Lungan *et al.* proposed multiple model hybrid methods to consider different models for different text structures [4]. By combining a one-vs-one (OVO) strategy with the Support Vector Machine (SVM),

Liu *et al.* proposed an optimized method to classify multiple emotions [5].

At the same time, traditional classification methods based on machine learning have also encountered many difficulties, especially the long learning time duration and the lack of computing power. For example, SVM requires exponential learning time and memory as the volume of data increases [6]. Due to the complexity of the text itself, there are problems in the corpus, such as the lack of universality and the no uniqueness of evaluation criteria.

Meanwhile, the classification method based on deep learning has been widely used because of its excellent extraction ability for the text features [7]. In the modern economic and social era, where the volume of data becomes large and the number of financial texts is growing rapidly, it may become a heavy task to improve the efficiency of classification and digest a large amount of financial professional texts published on the Internet. So, Kanung-sukkasem *et al.* introduced a topic model based on latent Dirichlet allocation (LDA) to discover features from news articles and financial time series. The financial LDA is applied in data mining for financial time series prediction and gets better results than the common LDA [8]. Shi *et al.* presented DeepClue, a system built to bridge text-based deep learning models and end users through visually interpreting the key factors learned in the stock price prediction model. The DeepClue predicted the stock price with financial news and company-related tweets from social media [9].

Now the amount of information on the Internet is huge, it is a common phenomenon that the financial text contents are uneven. For the information quality of financial texts is not uniform, or text information is not complete, it becomes extremely crucial to consider extracting features from part content of each article. However, according to the majority of current deep learning methods, the entire content of an article is taken as the training set and the testing set of classification, leading to less attention to the part of each financial text. In particular, so far, few studies using part of each article for the classification of financial text are developed. Thus, a new method proposed by this paper will fill in this gap. A new neural network for the text classification called "All Dataset based on CharCNN and GRU (AD-CharCGNN)". The network extracts features in the article parts via convolution operations and GRU, and classifies features via fully connected layer and softmax functions. The experiments confirm that the network works effectively with high accuracy. This network which takes full advantage of the text part is more suitable for practical application and fills in the gap about classification based on part content of each financial text.

The innovations and contributions of this paper are shown as below:

- In this paper, a text classification based on partial text is proposed. At present, on the Internet, it is a common phenomenon that the information quality of financial texts is not uniform, and text information is not complete. So, the text classification based on the partial text can simulate the incomplete information scene, and that means it is more suitable for real situations. Compared with the text classification of the whole text, the text classification of partial text is more difficult and challenging.

- In this paper, a neural network AD-CharCGNN combined charCNN and GRU is proposed. That means the AD-CharCGNN can obtain information from time domain and spatial domain. The AD-CharCGNN is based on character level. It is a frequent phenomenon that texts contain Chinese, English, numbers, or special characters. The network can handle all above characters, and character level network can read data set directly without processing it by deleting stopwords.

- In this paper, a Chinese data set in the financial field is built. Crawler technology was used to grab 65,000 pieces of text as the dataset from two authoritative financial websites in China. Compared with English text classification, Chinese text classification is more difficult because of the complexity of Chinese. At the same time, the text classification in this paper is a sub-class classification within the same professional field data set. As we all know, the differences among sub-classes in the same field are smaller than the differences among the large classes in different fields. This does make the text classification difficult, but it is beneficial to individuals or institutions in the financial field.

II. RELATED WORK
A. DEEP LEARNING MODEL FOR TEXT CLASSIFICATION
With the development of deep learning, the related model of text classification is also emerging. It is well known that the Convolutional Neural Networks (CNN) model has made outstanding achievements in image processing [10], [11], target detection [12], [13], and speech recognition [14], [15]. TextCNN algorithm proposed by Yoon Kim applied CNN model to text classification [16]. Using convolutional calculations, textCNN can obtain the local features and extract the key information similar to n-gram from a sentence. In spite of the outstanding performance in text classification, CNN's fixed filter size prevents CNN from capturing the full range context of the article. Hence, recurrent neural networks (RNN) have become one of the most popular architectures used in NLP problems because their recurrent structure is very suitable to process the variable length text [17].

The architectures of deep learning are diverse in text classification, but there is also some common ground among them. For example, most of these involve preprocessing the data sets and cleaning the text content via text segmentation and stopwords deleting. The feature words are mapped into a high dimensional spatial vector model, with word frequency as an important indicator of text classification. The importance of feature words is expressed through characteristic weight calculation, which means the whole text is mapped into a matrix to be processed by CNN.

These above models are classified at the word level. Inspired by the pixel level in the computer vision field, the Yann LeCun team proposed a model that is based entirely on charCNN to classify text [18]. The charCNN retrains the neural network from a character perspective. Experiments show that when the training set is large enough, the convolution network can achieve excellent results. At the same time, the network does not need the information on the word level, nor the grammar structure of the language. The charCNN can be applied to text classification of different languages because any language is made up of characters.

B. LSTM AND GRU METHOD

Long Short-Term Memory (LSTM) network, which is better at processing sequence than general RNN, can take the full context of the sentence into account. LSTM can be applied not only in text classification [19], but also in many scenes such as image segmentation [20], speech recognition [21], trajectory prediction [22] and so on. An example of the LSTM structure shown in Fig. 1 describes that the results generated by the last word are sent into a fully connected layer and to be classified by softmax function.
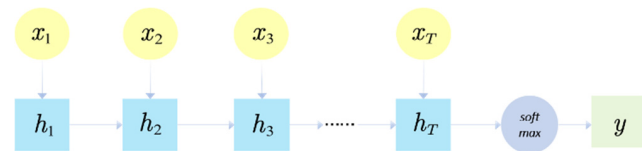

FIGURE 1. RNN with LSTM [17].

Bidirectional LSTM (BiLSTM) is a variant of LSTM for many applications, such as emotional analysis [23], entity recognition [24], relationship extraction [25], and so on. Liu *et al.* unified BiLSTM, attention mechanisms, and convolution layers into a network [26]. The network forms a two-way of long-term storage which captures both local characteristics of phrases and the global semantics of sentences. The text classification BiLSTM is as shown in Fig. 2. Firstly, the texts are mapped to vectors in the embedding layer, and then, features of vectors are extracted in the two-way LSTM layer to generate the last sequence. Finally, the last sequence will be classified in the fully connected layer with a softmax function.

Gated Recurrent Unit (GRU) is another variant of LSTM whose applications are also quite extensive [27]. Zhao *et al.* applied GRU to monitor machine health [28], Yuan *et al.* applied GRU to speech recognition [29], and Zhang *et al.* combined RNN with GRU to identify Chinese characters [30]. There are three types of gates in the LSTM: the forgetting gate, the inputting gate, and the outputting gate. GRU combines the forgetting gate and the inputting gate to form a separate "update gate". Owning exists update gate and reset gate only, the computational complexity of GRU is simpler than that of LSTM. The GRU architecture is shown in Fig. 3.
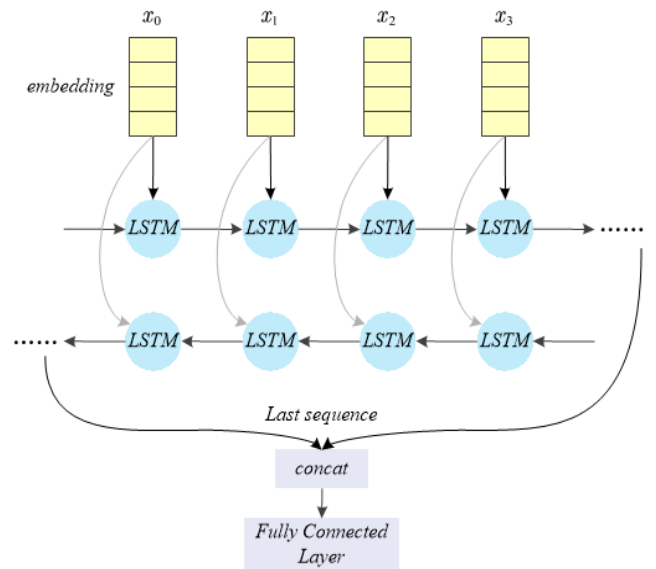

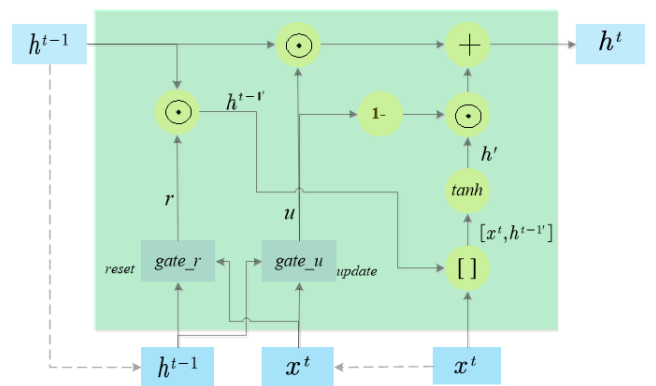**FIGURE 2.** The BiLSTM used in the text classification.

FIGURE 3. The GRU architecture.

The *gate_u* and *gate_r* in Fig. 3 are update gate and reset gate respectively. The h^{t-1} is the hidden state from the previous time and h^t is the hidden state of the current time. The x^t is the input of the current time. The degree of state information updating from the last time to the current time is controlled by the update gate. The reset gate determines how much state information last time is ignored.

So, the forward propagation and update state of GRU are as follows:

The state of *gate_u* is obtained by h^{t-1} and x^t:

$$u = \sigma(W_u \cdot [x^t, h^{t-1}]) \tag{1}$$

where [ ] indicates that two vectors are connected, W_u is a parameter that needs to be learned and σ is a sigmoid function. The formula converts the input data to a gate state value in the range of 0-1.

The closer u is to 1, the more ability to memory the *gate_u* owns. The closer u is to 0, the more ability to forget the *gate_u* owns.

The state of *gate_r* is also obtained by h^{t-1} and x^t:

$$r = \sigma(W_r \cdot [x^t, h^{t-1}]) \qquad (2)$$

where the meaning of operators is the same as in formula (1) and W_r is also a parameter that needs to be learned. $h^{t-1'}$ represents the remain information of h^{t-1} after reset:

$$h^{t-1'} = h^{t-1} \odot r \qquad (3)$$

where $\odot$ is the multiplication of corresponding elements between the two matrices.

The new proposed h' represents the hidden memory information of current input x^t. By tanh activation function, the value of h' can be reduced to a range of -1 to 1, which is:

$$h' = tanh(W_{r'} \cdot [x^t, h^{t-1'}]) \qquad (4)$$

where $W_{r'}$ is also a parameter that needs to be learned.

Update stage. Using the previously obtained state u, the update formula is as follows:

$$h^t = u \odot h^{t-1} + (1 - u) \odot h' \qquad (5)$$

where $+$ represents the add operation of vectors. The operation $u \odot h^{t-1}$ implements the selective forgetting of the previously hidden state information, that is, it can forget some unimportant information of h^{t-1}. The operation $(1 - u) \odot h'$ implements the selective memorizing of the current hidden state information, where it can forget some unimportant information of h'. Combined with the above, the calculation of this step is to give the remain information of h^{t-1} and h' to the currently hidden state h^t.

In summary, as is shown in Fig. 4, CNN is an unbiased model that can only obtain the most significant features of input texts in the spatial domain, whereas RNN is a bias model that can only describe the output of continuous state in the time domain. Different from a single CNN or RNN, the network proposed in this paper, AD-charCGNN, combines charCNN with RNN containing GRU and classifies financial professional texts in both spatial and time domains. In addition, the model does not require data cleaning for the preprocessing of the dataset. Financial texts from the Internet may contain Chinese, English, numbers, or special characters, which also means that AD-charCGNN should be more suitable for the actual scene.

III. THE AD-CharCGNN NETWORK

A. THE DESIGN OF NETWORK

The network architecture of AD-charCGNN is shown in Fig. 5. A professional Chinese financial text dataset is employed in this network. In the text processing stage, characters only needed to be classified are read into the vocabulary in charCNN, whereas AD-charCGNN reads all characters in the dataset and fills them into the character level vocabulary. Characters are converted into numbers in the vocabulary. This step makes it easy to map texts into high dimensional spatial vectors.
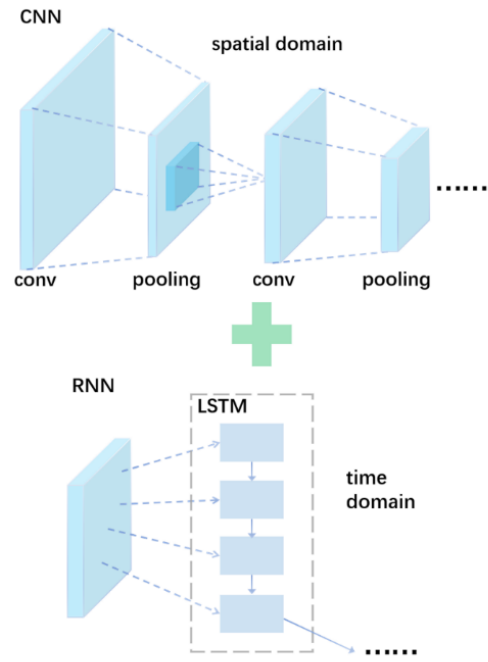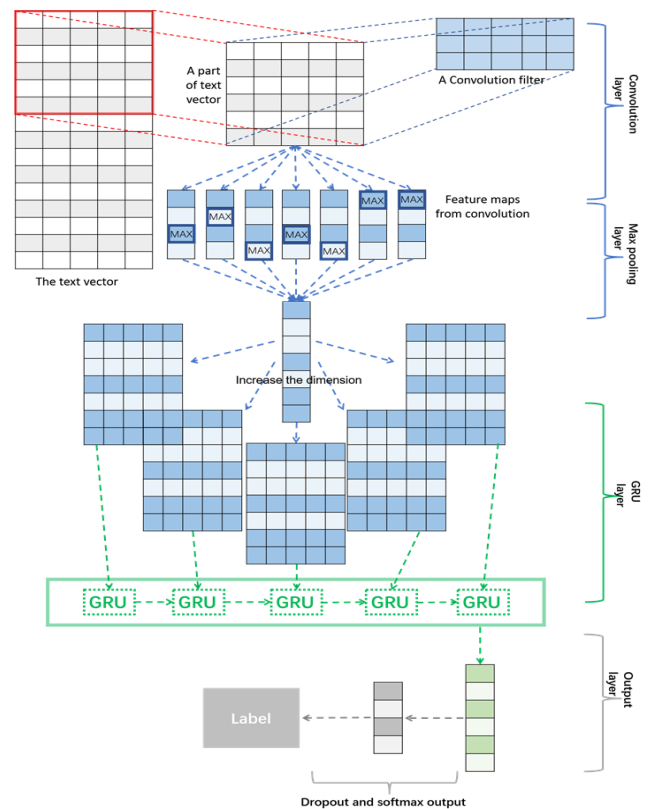

FIGURE 4. CNN and RNN model.

FIGURE 5. The AD-charCGNN network. The example is mapped to 5-dimensional space.

After the text processing stage, a part of each text is loaded into the network, which is the process from text vector to a part of the text vector in Fig. 5. The features of the text are obtained through a convolutional neural network. The width
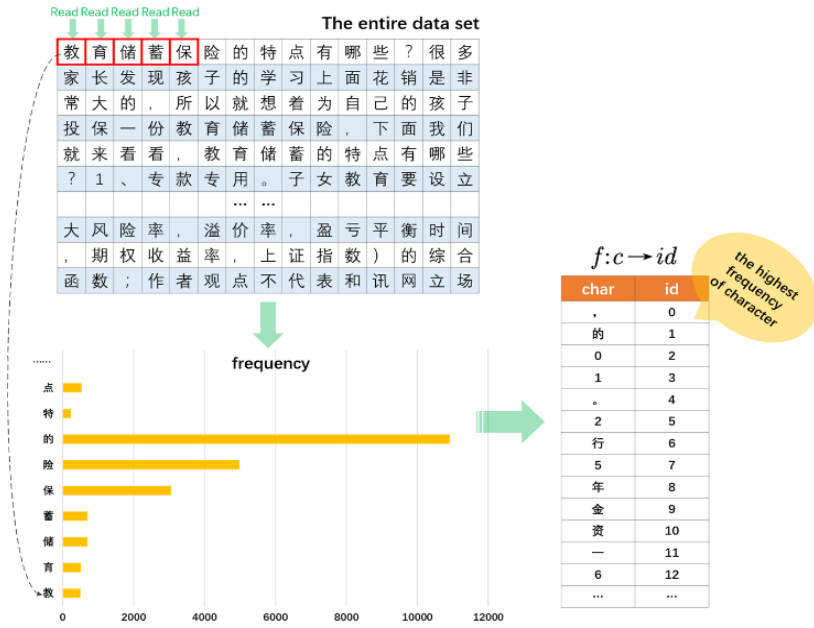
FIGURE 6. The mapping relation between character and number.

of the convolution filter should be the same as the width of the text vector. To be specific, the example in Fig. 5 is a 3*5 size convolution filter, and the filter acts on a 6*5 size text vector. For the one-dimensional convolution, the result of convolution is a feature map of 4*1 size. Multiple feature maps are gotten from multiple filters in the same way. Next, the max-pooling layer is used to obtain a significant feature of the part text. The feature map in Fig. 5 marked "MAX" is the maximum value of the map. The max-pooling layer can automatically determine which feature in the text classification process plays a more crucial role. Then, the GRU structure is used to obtain important contextual information. Finally, the classification task is completed through the softmax func

B. DATA PROCESSING

1) TEXT PROCESSING

According to the above, in order to achieve the part text classification through the AD-charCGNN network, the first step is to process the text and map it into high dimensional space in the form of vector.

All texts in the entire dataset are read character by character (as is shown in Fig. 6), including Chinese characters, English characters, punctuation symbols, special characters, and so on. Then the frequency of each character appeared would be counted. The mapping relationship based on the frequency would be formed. For example, in Fig. 6, the character "," is the most frequently occurring in the dataset, so its *id* is the smallest. The higher the frequency of the character, the smaller the number corresponding to the *id*. The mapping relation f between character c and number id is:

$$f : c \rightarrow id \qquad (6)$$

where $\rightarrow$ is a map symbol, that means the transformation from character to number. The *id* is also used as an index to establish the vocabulary.

The texts in the dataset are read again, and the characters of each text are converted into numbers according to the mapping relationship f, as follows:

$$A_n = [f(c_{n1}), f(c_{n2}), f(c_{n3}), \ldots, f(c_{nm})] \qquad (7)$$

where A_n is the nth text and c_{nm} is the mth character of the nth text.

The text number list A_n is then processed by the embedding layer. Embedding is a transformation method in NLP. It is well known that one-hot encoding is a common encoding transformation method, but the matrix obtained by one-hot is a sparse matrix, which will take up a lot of resources when the encoding content is too much. Embedding turns a sparse matrix into a dense matrix through some linear functions. This dense matrix uses some features to represent all characters and turns independent vectors into relational vectors with internal connections. Therefore, the embedding layer converts integers in A_n to high dimensional relational vectors of fixed size. In this paper, the financial text classification only intercepts a part of each text as the input A of the convolution layer:

$$A = \hat{A}_1 \oplus \hat{A}_2 \oplus \hat{A}_3 \oplus \ldots \oplus \hat{A}_n \qquad (8)$$

where $\hat{A}_n$ denotes the part text vector of the nth article in high dimension and $\oplus$ is the vector connector.

2) CONVOLUTION CAPTURES LOCAL INFORMATION

Convolution operation can not only capture the local information of the vectors but also reduce the dimension of the
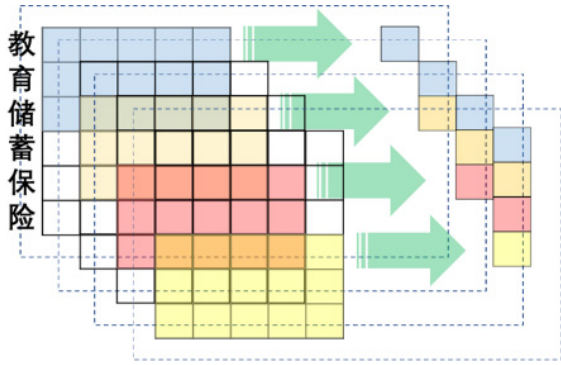
FIGURE 7. The one-dimensional convolution. Each character corresponds to a line, that is, a vector. As is shown in the figure, the color regions on the left represent the convolution templates. It is a convolution template of 3 × 5 acts on a vector of 6 × 5. The result is a vector of 4 × 1 after the one-dimensional convolution operations.

vectors and the computational cost of the model. The convolution layer applies a one-dimensional convolution core. As is shown in Fig. 7, The width of the convolution filter should be the same as the width of the text vector. Thus, a one-dimensional convolution core is used to convolute by row and detect features at different locations. The output after the convolution is:

$$A^* = F \otimes A + b \tag{9}$$

where F is the convolution template, b is the offset, and $\otimes$ is the convolution operation. Multiple convolutions outputs A^* can be generated via multiple convolutional core filters to capture more feature information. The output after convolution will be the input of the max-pooling layer. If the number of convolution cores is, the output T after pooling is:

$$T = \max\left(A_1^*\right) \oplus \max\left(A_2^*\right) \oplus \max\left(A_3^*\right) \oplus \ldots \oplus \max\left(A_l^*\right) \tag{10}$$

At this time, T is a feature vector that passed through convolution kernel l and is connected after max pooling. It is also an input of the gated recurrent unit in the next RNN.

3) INTRODUCTION OF GRU METHOD
After the convolution layer, the context feature information is read by the GRU to make dynamic modeling. The GRU hidden state is activated by the previous state in a certain period time. Therefore, among the two gates in the GRU, the degree of state information updating from the last time to the current time is controlled by the update gate. The reset gate determines how much state information the last time is written to the current candidate set.

Meanwhile, the GRU also contains three kinds of values, that are input value, output value, and candidate set. The particular GRU architecture about AD-charCGNN is shown in Fig. 8.

The T_t is defined as the input T at time t, and the calculation for updating $gate\_u_t$ and resetting $gate\_r_t$ are as follows:

$$u_t = \sigma\left((W_{uh} \cdot h_{t-1} \oplus W_{ut} \cdot T_t) + b_u\right) \tag{11}$$

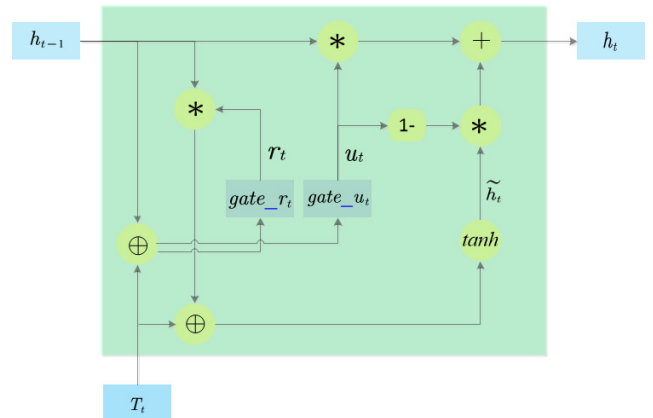$$r_t = \sigma\left((W_{rh} \cdot h_{t-1} \oplus W_{rt} \cdot T_t) + b_r\right) \tag{12}$$

FIGURE 8. The concrete representation of the GRU architecture.

where u_t and r_t are the state of $gate\_u_t$ and $gate\_r_t$ respectively. σ represents the sigmoid function. W_{uh}, W_{ut}, W_{rh}, and W_{rt} are parameters to be learned, h_{t-1} is the state passed from the previous time, b_u and b_r are offsets of update gate and reset gate respectively, and $\oplus$ is still a vector connector.

Candidate set $\tilde{h}_t$ of time t can be calculated according to $gate\_r_t$:

$$\tilde{h}_t = \tanh\left\{W_{\tilde{h}} \cdot ((r_t * h_{t-1}) \oplus T_t) + b_{\tilde{h}}\right\} \tag{13}$$

The activation function uses the tanh function. $W_{\tilde{h}}$ is the parameter that the candidate set needs to learn, $b_{\tilde{h}}$ is the offset of the candidate set, and the symbol $*$ represents the product of the matrix.

At the same time, h_t will be updated:

$$h_t = (1 - u_t) * \tilde{h}_t + u_t * h_{t-1} \tag{14}$$

Finally, the output Y_t of the output layer at time t is:

$$Y_t = \sigma\left(W_y \cdot h_t + b_y\right) \tag{15}$$

where W_y is the parameter to be learned and b_y is the offset in the output layer.

In this AD-charCGNN model, the last output sequence of GRU is selected as the output Y. Y is activated by dropout and relu function to obtain Y^*, which can be used as input information of the output layer.

4) CLASSIFIER FOR CLASSIFICATION
After the previous part of the network processing, the feature vectors need to be classified by classifiers to predict the classification probability of financial texts. The classification function selected is the softmax function, which can convert the output values of classification into relative probabilities and predict the probability of each class. The calculation formula is as follows:

$$y = softmax(Y^*) \tag{16}$$

Finally, the maximum probability item in y is selected as the predictive classification label for the final output.

IV. EXPERIMENTAL COMPARISON AND ANALYSIS

A. THE EFFECTIVENESS EXPERIMENT OF THE AD-CHARCGNN NETWORK

As we all know, the deep learning model will lead to different results due to various reasons such as hardware environment, parameter adjustment, and so on. Therefore, first of all, in order to confirm the effectiveness of the model proposed in this paper, and also to ensure that other state-of-art comparison models can achieve the optimal effect, we conducted a comparison test with some public data sets, and compared with the optimal accuracy of the original reference of the comparison model.

The public dataset used in the effectiveness experiment is shown below:

(1) **Reuters-21578:** It is a corpus which is often used for text classification or other related researches. Documents in Reuters-21578 were marked up with SGML tags, and a corresponding SGML DTD was produced. It is available from David D. Lewis' professional home page, currently: http://www.research.att.com/~lewis.

(2) **MR:** Movie Review Dataset [31]. This dataset contains movie reviews and their associated binary mood polarity labels. No more than 30 reviews are allowed for any given movie throughout the series, as reviews of the same film often have a relevant rating. The overall distribution of labels is balanced.

(3) **SST:** Stanford Sentiment Treebank Dataset [32]. The dataset was published by the NLP group at Stanford University. It is a standard emotion dataset, mainly used for emotion classification, in which each node of the sentence analysis tree has fine-grained emotion annotation.

The state-of-art comparison models used in the effectiveness experiment is shown below:

(1)**CNN:** Convolutional Neural Network [16]. The model where all words are randomly initialized and then modified during training.

(2)**DCNN:** Dynamic Convolutional Neural Network [33]. This network can use dynamic k-max pooling, a global pooling operation over linear sequences and can also handle input sentences of varying lengths.

(3)**RNN:** Recurrent Neural Network [17]. The multitask learning framework is used to jointly learn across multiple related tasks.

(4) **LSTM:** The LSTM model uses the last hidden state as the representation of the whole text [34].

(5)**Bi-LSTM:** A bi-directional LSTM, commonly used in text classification [34].

(6)**fastText:** A simple and efficient text classification method [35], which treats the average of word/n-grams embeddings as document embeddings, then feeds document embeddings into a linear classifier [34].

(7)**TextGCN:** Text Graph Convolutional Network [34]. It is initialized with one-hot representation for word and document, it then jointly learns the embeddings for both words

TABLE 1. Comparison of effectiveness experimental results.

Model	Reuters	MR	SST-1	SST-2
CNN	—	76.10	45.00	82.70
DCNN	—	—	48.50	86.80
RNN	—	—	**49.60**	87.90
LSTM	93.68	75.06	—	—
Bi-LSTM	96.31	77.68	—	—
fastText	96.13	75.14	—	—
TextGCN	**97.07**	76.74	—	—
DBN	86.88	—	—	—
charCGNN	95.31	**77.82**	47.44	**88.06**

and documents, as supervised by the known class labels for documents.

(8) **DBN:** Deep Belief Network [36]. After the feature extraction with DBN, softmax regression is employed to classify the text in the learned feature space.

With accuracy (%) as the experiment standard, the comparison results are shown in Table 1.

For the data set Reuters, the network TextGCN has the best effect, with an accuracy rate of 97.07%, while the network proposed in this paper has an accuracy rate of 96.31%, which is 1.76% lower than TextGCN. We can analyze the reasons according to the data set. In the data set Reuters, the text is divided into 8 classes, but the number of texts in each class is unbalanced. For example, the amount of texts labeled "grain" is fewer than 100. However, the network proposed in this paper is complex and its structure is deep, which inevitably leads to the poor learning of the network when the amount of data is small. This is a common shortcoming of complex neural networks.

The same problem shows up on SST. SST is an emotion data set. SST-1 means that the data set is divided into 5 classes, and SST-2 means that the data set is divided into 2 classes. As is shown in Table 1, the accuracy of charCGNN in SST-1 is 47.44%, which is 2.16% lower than RNN. This result is as we expected. The reason is still the small amount of data. When SST is divided into two classes, both positive and negative labels have sufficient data. It can be seen from the Table 1 that the accuracy of charCGNN in SST-2 is 88.06%, higher than other models.

As for the data set MR, the data set is also divided into positive and negative labels. The labels are evenly distributed and the data is enough for the network to learn. Therefore, the accuracy of charCGNN is 77.82%, which is also higher than other models. When the data is insufficient, we do not deny that the network does have incomplete learning. But when there is enough data, the network can perform well.

To sum up, the network proposed in this paper is valid. Then, this effective network is used to classify financial texts in more detail.

B. CONSTRUCTION AND DIVISION OF FINANCIAL DATASET

As we all know, in the data set, the differences between large classes in the different domains are obvious, which makes it easy to classify texts. However, the classification of small subclasses in the same domain is barely mentioned. On the one hand, the differences among subclasses in the same field are smaller than the differences among the large classes in different fields, so it is not easy to classify. On the other hand, most of the text classification at present is aimed at the public data sets of the large classes, and few people focus on the classification within a specific subject area. Subcategory text classification in a specific subject area can help people who are in the area to understand different kinds of information. In the case of the financial field, subcategory text classification can help financial individuals or financial institutions master more detailed news and make the right decisions. Especially for professional financial experts, after classification, detailed and effective financial texts can master the current advanced research and possible future directions, and comprehensively understand the financial information in network.

In view of the fact that the Internet does not contain publicly available financial datasets, financial news of 10 subclasses is crawled from SouthMoney and Hexun as the dataset. The website of SouthMoney (http://www.southmoney.com) is a well-known comprehensive financial and economic website in China. It covers the financial field with its authoritative industry analysis and multidirectional information. The site has 30 million users and is growing by 10,000 a day. The website of Hexun (http://www.hexun.com) is the first vertical website of financial information in China, which is the representative of the professional, high-end, and quality financial website. It works with lots of banking institutions, fund companies, and media. It has also worked with Thomson Reuters, which is the world's largest provider of financial information data and analytics.

The financial dataset contains 10 subclasses:

(1) **Insurance:** It is an important pillar of the financial system and the social safety net. It contains insurance industry dynamics, as well as related views and comments.

(2) **Stock Market:** It is one of the main long-term credit instruments in the capital market and an indispensable part of the financial area. It includes the expert interpretation of the stock market, also includes the individual stock, state shares, and other kinds of news about the stock.

(3)**Companies:** It includes the company's hot news, the company's operations, and the impact of national policies on the company.

(4) **Funds:** It includes fund information, fund analysis, fund knowledge, fund assessment, and other fund news.

(5) **Futures:** It includes oil, coal, gold, and other futures trend analysis news, also includes all kinds of futures operation strategy, market preview, and other relevant news.

(6) **Automobiles:** It includes dynamic news of all kinds of auto brands, such as the release of new cars, the sales volume of cars, and so on.

(7) **Foreign Exchange:** It includes the currency situation of each country, including but not limited to RMB, us dollar, pound sterling, euro, yen, and so on.

(8) **Trust:** It is a crucial part of the modern financial system. It contains the information of trust products, the dynamics of trust industry, the comments of the trust study, and other relevant news.

(9) **Banks:** It includes the deposit interest rate of each bank, loan analysis, industry policy, and other relevant contents.

(10) **Bonds:** It includes all kinds of news related to bond buybacks, bond knowledge, bond prices, and so on.

The overall distribution of labels is balanced. There are 6500 texts of each class, for a total of 65,000. The specific division of the dataset is shown in Table 2.

TABLE 2. Dataset division.

	Train	Val	Test
Number/class	5000	500	1000
Total Number	50000	5000	10000

C. THE SETTING OF FINANCIALLY EXPERIMENTAL PARAMETERS

The dataset used is a professional financial dataset mentioned above in this financial experiment. According to the web crawler results, the financial texts that exist on the Internet are varied in length. The texts can reach up to 1000 words, but at least only about 100 words. First of all, in order to distinguish from short text (100 to 200 words) classification, 250/300/ 350/400 words of text parts were selected for the test. Test results show that text classification accuracy of 400 words is the highest, and the accuracy of 250 words is the lowest. Therefore, the following parameters debugging are the 250 words texts, which are made as a part of each article in the experiment. The number "0" is filled in the absent word place when the length of the article less than 250 words.

The parameters of the AD-charCGNN network are set as Table 3 after several parameter tunings.

TABLE 3. AD-CharCGNN network parameter settings.

Parameter	Value
Length of Text Part	250
Kernel Size	3
Dropout Rate	0.7
Learning Rate	0.001
Number of Filters	128
Dimension of Embedding	64
Dimension of Hidden Neurons	128
Max of Total Batch	1000

The "Max of Total Batch" is used to prevent overfitting. If the accuracy of the validation set is not improved for a long time, that is, it is not improved more than 1000 rounds, then the training is terminated in advance to prevent overfitting.

The neural network uses the cross-entropy error function as the loss function. The cross entropy is often used as a loss function in the classification of neural networks. The function calculates the cross entropy between the predicted category probability and the real category.

D. COMPARISON OF FINANCIALLY EXPERIMENTAL RESULTS AND ANALYSIS

The dataset used is the Chinese financial text dataset. The inputting data are also text parts. The baselines for comparison experiments are as follows:

(1) **LR:** Logistic Regression. Logistic Regression measures the relationship between variables which dependent on category and one or more independent variables.

(2) **NB:** Naive Bayes. A classification technique based on the Bayesian theorem assumes that a particular feature in a class has nothing to do with other existing features.

(3) **RF:** Random Forest. It is an integrated learning method based on Bagging.

(4) **Xgboost:** It is another tree-based integration model. It can upgrade weak learners to strong learners.

With Accuracy (ACC) as the experiment standard, the comparison results are shown in Table 4.

In Table 4, "Count Vectors" represents that the feature engineering for text processing is counting vectors. "Word Level TF-IDF" represents that the feature engineering is a word-level frequency. "N-Gram Vectors" represents that the feature engineering is an n-gram level word frequency. As can be seen from Table 4, it proves that the neural network proposed in this paper can effectively classify subclasses in the financial field. Compared with traditional machine learning methods, AD-charCGNN performs well with an accuracy of 96.45%, which is 1.33% higher than charCNN.

No matter AD-charCGNN or charCNN, compared with the traditional algorithm, it can capture more key features through convolution, so as to achieve higher accuracy than the traditional algorithm. In particular, when the input content is a part of the text, the traditional algorithm cannot extract crucial information in the case of incomplete text. We know that charCNN and AD-charCGNN are both neural network models. However, the essence of charCNN is still a convolutional neural network, whose filter window is fixed. The fixed window limits the ability to capture full context information. AD-charCGNN combines CNN and GRU to perform convolution operation in the spatial domain, and also take the context belonging to the time domain into consideration. In other words, the neural network proposed in this paper convolves the input text with a convolution template to extract important features, which is an operation belonging to the spatial domain. At the same time, because each text has input before and after the current time, the GRU treats the context of the text as a time series, which is an operation belonging to the time domain. So, the network can obtain information in a relatively complete way and achieve a better classification effect.

Then, the next analysis is carried out by visualizing the accuracy value and loss value. As the number of iterations increases, the accuracy of the AD-charCGNN model training set is shown in Fig. 9. The loss change of the training set is shown in Fig. 10. The accuracy of the validation set is shown
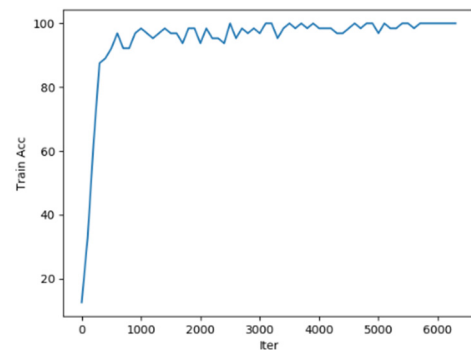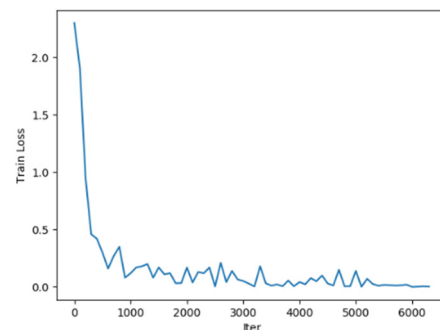
FIGURE 9. The accuracy of the training set.

TABLE 4. Comparison of financially experimental results.

Algorithm	ACC(%)
LR, Count Vectors	70.57
LR, Word-Level TF-IDF	64.02
LR, N-Gram Vectors	37.67
NB, Count Vectors	73.34
NB, Word-Level TF-IDF	63.90
NB, N-Gram Vectors	37.66
RF, Count Vectors	59.13
RF, Word-Level TF-IDF	58.78
Xgboost, Count Vectors	54.07
Xgboost, Word-Level TF-IDF	54.06
charCNN	95.12
AD-charCGNN	**96.45**

FIGURE 10. The loss changes of the training set.

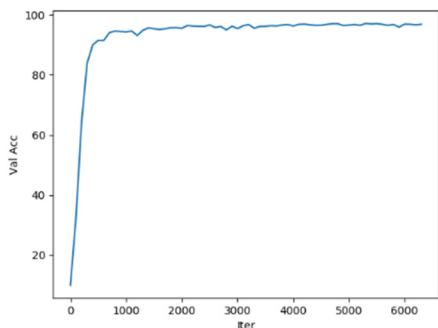in Fig. 11. The loss change of the validation set is shown in Fig. 12.

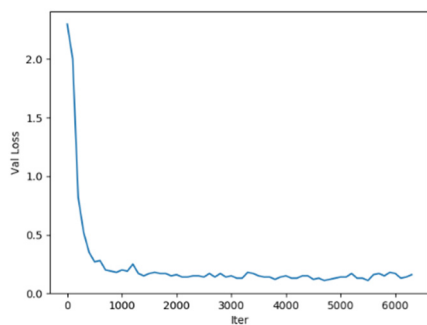**FIGURE 11.** The accuracy of the validation set.

FIGURE 12. The loss changes of the validation set.

FIGURE 13. Category assessment.

TABLE 5. Category assessment.

	Precision	Recall	F1-score
Insurance	0.99	0.99	0.99
Stock Market	1.00	0.99	0.99
Companies	0.90	0.91	0.91
Funds	0.97	0.95	0.96
Futures	0.97	0.98	0.97
Automobiles	0.97	0.94	0.95
Foreign Exchange	0.98	0.99	0.98
Trust	0.99	0.99	0.99
Banks	0.96	0.93	0.95
Bonds	0.92	0.98	0.95

It can be found that there is no overfitting in this model. As is known to all, if the model can obtain superior fitting on the training data, but it cannot fit the data well on the data set outside the training data, then the phenomenon of overfitting can be considered. As is shown from Fig. 9 to Fig. 12, it can be observed that the accuracy of the training set and the validation set both fluctuate at about 96%, and the loss value fluctuates around 0.14. Compared with the training set, the accuracy value and the loss value of the validation set are smoother and more stable. So, it is certain that the model has not been overfitted.

At the same time, the experiment also uses precision, recall, and F1-score to evaluate the performance of AD-charCGNN in each class [37], as is shown in Table 5 and Fig. 13.

As is shown in Fig. 13, the network performs well in the classes of Insurance, Stock Market, Foreign Exchange, and Trust. The values of subclass Companies are relatively low. We can analyze the reasons according to the specific news content. Take the news of Stock Market as an example, the focus of this news is very clear, that is to say, when the news is written from the perspective of Stock Market, the content is all around the stock, such as the rise and down of Stock Market. And the focus of Companies can sometimes be unclear. The news about a company, for example, will include the information about what has happened to the company's
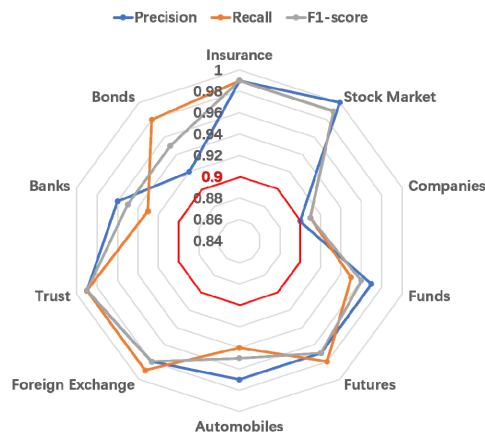
funds and whether the company has good credit with the bank. As a result, the Companies subclass may be mistakenly assigned to other subclasses. But on the whole, the precision, recall, and F1-score of these classes are all above 0.9, and the neural network proposed in this paper has achieved good results.

V. CONCLUSION

The classification of financial texts is an indispensable step to take advantage of financial news. But the quality of Internet financial texts is uneven. To classify incomplete information, the AD-charCGNN network is proposed.

Firstly, all characters in the financial texts are read into the network to build a character level vocabulary of the financial dataset created by ourselves. Secondly, part of the text which will be classified is mapped to a high-dimensional spatial vector based on the vocabulary. Then, the vectors are convoluted in the spatial domain to get the text local features, and next, the features are processed by the gated recurrent units to get the features contained time information. Finally, the features will be classified through softmax function to get the text classification results. The text classification algorithm which takes advantage of the text part is more suitable for practical application. For financial texts, which may contain characters of Chinese, English, digital and other types of characters,

the network makes it easier to process the various type of character text.

The dataset applied in the paper is based on the Chinese financial text. In the future, the dataset will be extended to other fields or other languages of professional text, which is available to build a cross-language text classification model.

REFERENCES

[1] M. Piškorec, N. Antulov-Fantulin, P. K. Novak, I. Mozetic, M. Grcar, I. Vodenska, and T. Šmuc, "Cohesiveness in financial news and its relation to market volatility," *Sci. Rep.*, vol. 4, p. 5038, May 2014.

[2] L. Zhou, H. Wang, L. Zhang, E. Chen, and J. Chen, "A novel knowledge network framework for financial news navigation," in *Web-Age Information Management*. Springer, 2014, pp. 723–727.

[3] S. Kalra and J. S. Prasad, "Efficacy of news sentiment for stock market prediction," in *Proc. Int. Conf. Mach. Learn., Big Data, Cloud Parallel Comput. (COMITCon)*, Faridabad, India, Feb. 2019, pp. 491–496.

[4] L. Zhang, L. Jiang, and c. Li, "A discriminative model selection approach and its application to text classification," *Neural Comput. Appl.*, vol. 31, nos. 2–3, pp. 1173–1187, Jul. 2017.

[5] Y. Liu, J.-W. Liu, and Z.-P. Fan, "A method for multi-class sentiment classification based on an improved one-vs-one (OVO) strategy and the support vector machine (SVM) algorithm," *Inf. Sci.*, vols. 394–395, pp. 38–52, Jul. 2017.

[6] Y. Chae, D.-H. Jeong, and T. Kim, "Experimental study of time series-based dataset selection for effective text classification," in *Proc. Int. Conf. Knowl. Smart Technol.*, Feb. 2017, pp. 354–358.

[7] J. Lu, C. X. Ma, and Y. R. Zhou, "Classification of multi-category text information under the framework of text-CRNN+attention," *Appl. Res. Comput.*, pp. 1–6, 2019, doi: 10.19734/j.issn.1001-3695.2018.12.0858.

[8] N. Kanungsukkasem and T. Leelanupab, "Financial latent Dirichlet allocation (FinLDA): Feature extraction in text and data mining for financial time series prediction," *IEEE Access*, vol. 7, pp. 71645–71664, 2019.

[9] L. Shi, Z. Teng, L. Wang, Y. Zhang, and A. Binder, "DeepClue: Visual interpretation of text-based deep stock prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 6, pp. 1094–1108, Jun. 2019.

[10] X. Yang, Y. Ye, X. Li, R. Y. K. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5408–5423, Sep. 2018.

[11] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li, "Hyperspectral image super-resolution using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 29–41, Nov. 2017.

[12] Y. Wang, Y. Zhang, H. Qu, and Q. Tian, "Target detection and recognition based on convolutional neural network for SAR image," in *Proc. 11th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2018, pp. 1–5.

[13] D. Zhao, H. Zhou, S. Rang, and X. Jia, "An adaptation of CNN for small target detection in the infrared," in *Proc. IGARSS-IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 669–672.

[14] J. Gu, Z. Wang, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, T. Chen, and J. Kuen, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.

[15] S. Ganapathy and V. Peddinti, "3-D CNN models for far-field multi-channel speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 5499–5503.

[16] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1746–1751.

[17] P. Liu, X. Qiu, and X. Huang, "Recurrent neural network for text classification with multi-task learning," 2016, *arXiv:1605.05101*. [Online]. Available: http://arxiv.org/abs/1605.05101

[18] X. Zhang, J. Zhao, and Y. LeCun, "Character-level convolutional networks for text classification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 649–657.

[19] L. Xiao, G. Wang, and Y. Zuo, "Research on patent text classification based on word2vec and LSTM," in *Proc. 11th Int. Symp. Comput. Intell. Design (ISCID)*, Hangzhou, China, Dec. 2018, pp. 71–74.

[20] K. Xie and Y. Wen, "LSTM-MA: A LSTM method with multi-modality and adjacency constraint for brain image segmentation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Taipei, Taiwan, Sep. 2019, pp. 240–244.

[21] Y. Zhang and X. Lu, "A speech recognition acoustic model based on LSTM-CTC," in *Proc. IEEE 18th Int. Conf. Commun. Technol. (ICCT)*, Chongqing, China, Oct. 2018, pp. 1052–1055.

[22] S. Dai, L. Li, and Z. Li, "Modeling vehicle interactions via modified LSTM models for trajectory prediction," *IEEE Access*, vol. 7, pp. 38287–38296, 2019.

[23] T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," *Expert Syst. Appl.*, vol. 72, pp. 221–230, Apr. 2017.

[24] L. Luo, Z. Yang, P. Yang, Y. Zhang, L. Wang, H. Lin, and J. Wang, "An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition," *Bioinformatics*, vol. 34, no. 8, pp. 1381–1388, Apr. 2018.

[25] S. Zheng, Y. Hao, D. Lu, H. Bao, J. Xu, H. Hao, and B. Xu, "Joint entity and relation extraction based on a hybrid neural network," *Neurocomputing*, vol. 257, pp. 59–66, Sep. 2017.

[26] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, Apr. 2019.

[27] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: http://arxiv.org/abs/1406.1078

[28] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.

[29] Y. Yuan, C. Tian, and X. Lu, "Auxiliary loss multimodal GRU model in audio-visual speech recognition," *IEEE Access*, vol. 6, pp. 5573–5583, 2018.

[30] X.-Y. Zhang, F. Yin, Y.-M. Zhang, C.-L. Liu, and Y. Bengio, "Drawing and recognizing chinese characters with recurrent neural network," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 849–862, Apr. 2018.

[31] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, Hum. Lang. Technol., Assoc. Comput. Linguistics*, Portland, OR, USA, Jun. 2011, pp. 142–150.

[32] R. Socher, J. Bauer, C. D. Manning, and A. Y. Ng, "Parsing with compositional vector grammars," in *Proc. EMNLP*, 2013, pp. 455–465.

[33] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," vol. 1, 2014, *arXiv:1404.2188*. [Online]. Available: http://arxiv.org/abs/1404.2188

[34] L. Yao, C. S. Mao, and Y. Luo, "Graph convolutional networks for text classification," in *Proc. 33rd AAAI Conf. Artif. Intell.*, 2018, pp. 7370–7377.

[35] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, vol. 2, 2017, pp. 427–431.

[36] M. Jiang, Y. Liang, X. Feng, X. Fan, Z. Pei, Y. Xue, and R. Guan, "Text classification based on deep belief network and softmax regression," *Neural Comput. Appl.*, vol. 29, no. 1, pp. 61–70, Jan. 2018.

[37] S. Yildirim, D. Jothimani, C. Kavaklioglu, and A. Basar, "Classification of 'hot news' for financial forecast using NLP techniques," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 4719–4722.

WENJIE ZHAO was born in Tangshan, Hebei, China. She received the B.E. degree from the College of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan, Shanxi, in 2014. She is currently pursuing the master's degree with the Shanghai University of Engineering Science, Shanghai, China. Her major research interest includes natural language processing.

GAOYU ZHANG is an Associate Professor with the Shanghai Lixin University of Accounting and Finance, China. His research interests include social computing, natural language understanding, and deep learning.
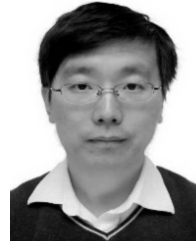
GEORGE YUAN received the B.Sc. degree in mathematics from Sichuan University, Chengdu, China, in 1986, and the Ph.D. degree in mathematics from Dalhousie University, Halifax, NS, Canada, in 1993.

He was a Postdoctoral Research Fellow with the Department of Mathematics, The University of Queensland, Brisbane, QLD, Australia, from January 1994 to September 1999. He was a Risk Manager with the Banks of Montreal Group, Toronto, Canada, from October 1999 to May 2001. He was then a Direct Manager with TXU Energy Trading Company, Dallas, USA, from May 2001 to December 2002. He was the Director of Risk Advisory Service with KPMG, from January 2003 to August 2018 and then the Director of Financial Advisory Service with Deloitte, from September 2008 to December 2012. In January 2013, He jointed Tongji University, Shanghai; the Shanghai Lixin University of Accounting and Finance, Shanghai; the Business School, Sun Yat-sen University, Guangzhou; and the Guizhou University of Finance and Economics, Guiyang, where he is currently a Chair Professor in fintech and financial engineering. His main research interests include fintech, financial engineering, financial mathematics, nonlinear analysis, game theory, and related applications to mathematical economics and optimization. He has published over 150 research articles in these areas. He successfully supervised over ten Ph.D. students and more than 100 M.Sc. students to completion with focus on the practice of financial engineering and financial technology in the industries.

JUN LIU is currently an Associate Professor with the Shanghai Lixin University of Accounting and Finance, China. His research interests include online teaching, learning science, and education technology.

HONGTAO SHAN is an Associate Professor with the Shanghai University of Engineering, China. Her research interests include natural language under-standing, digital control technology, and deep learning.

SHUYI ZHANG received the bachelor's degree in mathematics and the M.A. degree from the University of Science and Technology of China, Hefei, China, in 1990, and the Ph.D. degree in management from Southwest Jiaotong University, Chengdu, China, in 2003.

He served a large scale state-owned enterprise focused on railway civil engineering for 13 years. Since then, he has been working with the Zhejiang University of Science and Technology, Hangzhou, Shanghai Finance University, and the Shanghai Lixin University of Accounting and Finance. He has published 50+ articles in *International Entrepreneurship and Management Journal*, the *European Journal of International Management*, and so on, completed national and provincial research projects, and authored some book chapters also. His current research interests include technology innovation, entrepreneurial finance, with a focus on integrating global science network resources in the era of fintech and AI.

● ● ●