IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 12, 2021, accepted January 15, 2021, date of publication January 19, 2021, date of current version January 27, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3052937

Combination of Convolutional Neural Network and Gated Recurrent Unit for Aspect-Based Sentiment Analysis

NARISA ZHAO, HUAN GAO, XIN WEN, AND HUI LI
Institute of Systems Engineering, Dalian University of Technology, Dalian 116023, China

Corresponding author: Xin Wen (xinwenim@mail.dlut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61471083, and in part by the Dalian Science and Technology Innovation Foundation Project under Grant 2018J11CY009.

ABSTRACT Aspect-based sentiment analysis (ABSA) aims to identify views and sentiment polarities towards a given aspect in reviews. Compared with general sentiment analysis, ABSA can provide more detailed and complete information. Recently, ABSA has become an important task for natural language understanding and has attracted considerable attention from both academic and industry fields. The sentiment polarity of a sentence is not only decided by its content but also has a relatively significant correlation with the targeted aspect. For this reason, we propose a model for aspect-based sentiment analysis which is a combination of Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU), utilizing the local features generated by CNN and the long-term dependency learned by GRU. Extensive experiments have been conducted on datasets of hotels and cars, and results show that the proposed model achieves excellent performance in terms of aspect extraction and sentiment classification. Experiments also demonstrate the great domain expansion capability of the model.

INDEX TERMS Aspect-based sentiment analysis, online reviews, neural networks.

I. INTRODUCTION

The rapid development of e-commerce has brought about the accumulation of abundant consumer-generated reviews. These reviews are highly valuable in both economic and social aspects because we can extract the opinions of users by operating sentiment analysis on the product reviews. The opinions obtained can not only help potential consumers with purchase analysis but also provide the advantages and disadvantages of the goods to providers so that they can make further improvements in their products or services. Over the years, sentiment analysis [1] has gained increasing popularity in processing data from social media, such as online communities, blogs, Weibo and other online collaborative platforms.

Sentiment analysis, also named opinion mining [2], is an essential task in natural language processing and has attracted considerable attention both in academia and industry fields, particularly for identifying consumer satisfaction with products and services. Early sentiment analysis research [3] simply focused on the overall sentiment polarity of the given text. The work is established on the basis that there is just one general sentiment polarity in the text, which is equivalent to the coarse-grained sentiment analysis. However, most product reviews involve various aspects and the sentiment polarity towards each aspect may vary from each other. For example, the review 'Staff there were not friendly but the taste was good.' contains both positive and negative sentiments. Here, the sentiment polarity of the service aspect is negative, whereas the sentiment polarity of the food aspect is positive. This example reveals the necessity of conducting aspect-level sentiment analysis on reviews to provide more complete and detailed information.

Aspect-based sentiment analysis (ABSA) [4] refers to the task of identifying different aspects in a review and distinguishing the sentiment polarities towards each aspect separately to extract fine-grained information. The task of ABSA is of great research value since it enables consumers to evaluate the service or product from a comprehensive perspective so that they can obtain a more detailed and explicit understanding of it. Although current studies on neural networks have made significant progress and have performed notably well on many natural language processing tasks, such as machine translation [5] and text summarization [6], ABSA with neural networks still needs to be studied in depth. Moreover, most of the existing researches on ABSA are in English,

The associate editor coordinating the review of this manuscript and approving it for publication was Huiling Chen.

and we could not find many studies in Chinese on this topic.

In this work, we propose a combined Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) neural network model that utilizes local features generated by CNN as the input of GRU for ABSA on Chinese online review datasets of hotels and cars. We take word embeddings as the input of CNN, randomly set all values in a specific dimension as zeros after the convolution operation, and then use the encoded feature maps as the input of GRU. The long-term dependency learned by GRU can be seen as the sentence-level representations, and features that have stronger correlations with labels extracted by global max pooling and global average pooling operations are finally inputted into a fully connected layer, with a softmax layer serving as the output function. The proposed model makes full use of the CNN-extracted local features and long-term dependency of GRU, and gives a finer analysis on reviews to solve ABSA, showing excellent performance on hotel and automotive datasets.

The remainder of paper is organized as follows: Section II reviews related works on sentiment analysis. Section III illustrates our research task and presents the architecture of the proposed model. Section IV describes details about the setup and results of the experiments. The conclusion is drawn in section V.

II. RELATED WORK

A. NEURAL NETWORKS

Over the years, some classic deep learning algorithms have achieved remarkable success in computer vision and speech recognition fields [7], such as Recursive Neural Network (RNN) [8], Recurrent Neural Network (RNN) [9] and CNN [10]. Due to the good effect of deep learning methods, they have also received considerable attention from scholars in the field of natural language processing [11]. It has been shown that deep learning methods can automatically extract features from texts, and that they could perform better than traditional machine learning methods [12].

Since Mikolov presented a simple and effective way for the distributed representation of words [13], neural networks have greatly promoted the development of sentiment analysis. Kim [14] introduced an English text classification model which took preprocessed word vectors as the input and used CNN to implement the sentence-level classification task. Zhu *et al.* [15] used Long short-term Memory (LSTM) over sentiment analysis to model word sequences of online reviews. Ren *et al.* [16] used LSTM and combined the theme features to extract the features of Twitter short texts, which also made great progress in the experiment.

B. ASPECT-BASED SENTIMENT ANALYSIS

Sentiment analysis aims to classify a text as positive, negative and neutral according to the emotion tendency involved in the text [17]. General document-level sentiment analysis can only determine the overall sentiment polarity of it. However, the overall sentiment polarity towards a product or service can be inconsistent with the sentiments of some specific aspects. If we ignore the aspect extraction task, the accuracy of sentiment analysis could be lowered when multiple aspects are covered in reviews. ABSA is a type of fine-grained sentiment analysis, which is the task of identifying sentiment polarities towards a set of various aspects in texts. According to the SemEval 2014 Task 4 [18], a sentiment analysis task can be classified into two main categories: aspect-term sentiment analysis and aspect-category sentiment analysis.

Most of the existing studies on identifying the overall sentiment polarity of a text are coarse-grained sentiment analyses, which ignored aspects or entities in the text. Early on, with a growing number of sentiments lexicons [19], Mohammad [20] extracted the features of lexicons for sentiment analysis. Meng *et al.* [21] utilized sentiment lexicons on opinion pair recognition and matching, and subjective sentiment word detection tasks. These studies prove that the performance of sentiment analysis based on sentiment lexicons depends largely on the quality of features. Medhat *et al.* [22] treated sentiment analysis as the task of sentiment recognition and classification. Thus, they took product reviews as an example and divided the sentiment analysis process into four steps as follows: sentiment recognition, product property selection, sentiment classification and sentiment polarity identification. Xue *et al.* [23] built several models for fine-grained sentiment analysis, aiming at dealing with two subtasks that are aspect-category sentiment analysis (ACSA) and aspect-target sentiment analysis (ATSA). Motivated by the success of deep learning, Dong *et al.* [24] proposed an adaptive recursive neural network for target-related Twitter sentiment classification. Liang *et al.* [25] introduced a convolutional neural network for ABSA which can capture multi-layered sentiment information. Xiao and Zhou [26] designed a model called ASEGC which is based on the graph convolutional networks to better leverage edge information on ABSA task and achieved comparative performance.

Although RNN has made breakthrough improvements in text classification, it is still a challenge to infer sentiment polarity at a fine-grained level. A majority of researchers did not consider local features or deep-seated relations among features in texts when they implemented RNN in ABSA tasks, which resulted in comparatively lower accuracy for aspect extraction and sentiment analysis. To address this problem, this paper proposes a neural network model for ABSA, combining CNN and GRU. We utilize the local features extracted by CNN as the input of GRU, and then obtain long-term dependency learned by GRU. Finally, we perform sentiment analysis on Chinese online review datasets of hotels and cars.

III. ASPECT-BASED SENTIMENT ANALYSIS RESEARCH

A. TASK DESCRIPTION

This work aims to deeply mine valuable information in online reviews and analyze the sentiments towards various aspects of a certain product that consumers are concerned about.

Given a product review dataset $D = \{d_1, d_2, d_3, \ldots, d_N\}$, denote N as the number of reviews, d_i as the i-th review and assume review d_i involves several aspects with different sentiment polarities. We frame the ABSA task as two problems: (1) Extracting aspects $\{A_1, A_2, A_3, \ldots, A_M\}$ that consumers are concerned about from online reviews objectively and accurately. Here, M is the number of aspects, and A_j represents the j-th word that represents the corresponding aspect. (2) Performing a finer-grained sentiment analysis on aspects in reviews to obtain the sentiment polarities towards each aspect. S_i^j indicates the sentiment polarity towards aspect A_j in the i-th review, and $S_i^j \in \{-2, -1, 0, 1\}$, where $S_i^j = -2, S_i^j = -1, S_i^j = 0$, and $S_i^j = 1$ respectively represent that the sentiment polarity towards A_j is unmentioned, negative, neutral and positive. To sum up, the task of this paper is to determine $\left(S_1^1, S_1^2, \ldots, S_1^M; S_2^1, S_2^2, \ldots, S_2^M; \ldots; S_N^1, S_N^2, \ldots, S_N^M\right)$.

B. REPRESENTATION OF REVIEWS

The purpose of this paper is to identify sentiment polarities towards different aspects of Chinese product reviews. Therefore, text data should be represented with computer language for later process and analysis. We first use the word segment tool Jieba (https://github.com/fxsjy/jieba) for tokenization, then utilize the public stop words list of Baidu (http://www.baiduguide.com/baidu-stopwords/) to remove stop words in reviews. Finally, conduct unsupervised training on word embeddings with word2vec and transform each word into a word vector. Set the size of bag-of-words as Ψ; the dimensions of the word vector as $\boldsymbol{U}$ and the word vector matrix is $\boldsymbol{Q} \in \mathbb{R}^{\Psi \times U}$, then encode all of the words in reviews with row vectors of the word vector matrix. Pad or cut every review to make its length equal to a fixed value L, then each review d in the review set can be represented as a sequence as follows:

$$[w_1, w_2, w_3, \ldots, w_L] \tag{1}$$

$$\boldsymbol{E_i} = \boldsymbol{Q}[w_i], \quad \boldsymbol{E_i} \in \mathbb{R}^U \tag{2}$$

where $w_i (1 \leq i \leq \mathrm{L})$ denotes the i-th word in d, $\boldsymbol{E_i}$ is the word vector of w_i, and in this way, the review text d can be represented as the following word vector sequence:

$$\boldsymbol{E} = [E_1, E_2, E_3, \ldots, E_L], \quad \boldsymbol{E} \in \mathbb{R}^{L \times U} \tag{3}$$

C. MODEL ARCHITECTURE

CNN can extract local features and deep features from natural languages, while an RNN can process sequential input and solve long-term dependency problems. In this paper, we combine them for the ABSA of reviews. The architecture of the model is shown in Fig. 1. The model is composed of the following parts: convolutional layer, bidirectional GRU (BGRU) layer, pooling layer, concatenation layer, and fully connected layer with a softmax output.

1) CONVOLUTIONAL LAYER

Sharing weights among convolutional layers allows the model to remove some connections between layers in the network, which can help to reduce the calculation and avoid overfitting. In this paper, we pad the primary data so that the size of samples after the convolution operation would keep unchanged. After padding, conduct the convolution operation on sentence representation sequences. Denote the convolutional filters as $H \in \mathbb{R}^{h \times U}$ for each window size of 5, a feature c_i is generated from applying a convolutional kernel H to a window by:

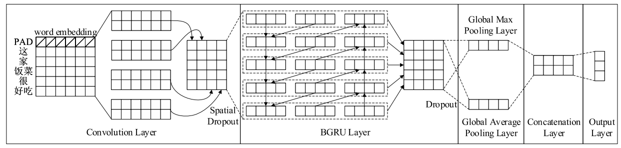

FIGURE 1. Model architecture for an example sentence.

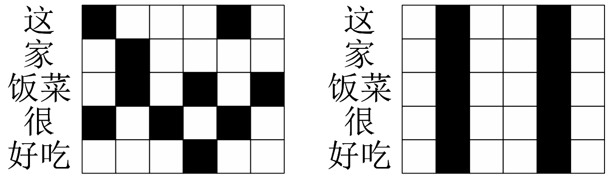

FIGURE 2. Distinction between SpatialDropout and Dropout (Left: Dropout, Right: SpatialDropout).

$$c_i = f\left(H \circ E_{i;i+h-1} + b\right) \tag{4}$$

where $E_{i;i+h-1} \in \mathbb{R}^{h \times U}$ are column vectors ranging from i to (i + h − 1), $b \epsilon \mathbb{R}$ is a bias, and an f is a nonlinear activation function. In our work, we use the ReLU activation function:

$$f(x) = max(0, x) \tag{5}$$

Applying convolutional kernels on the entire input to get a feature map:

$$\boldsymbol{C_i} = [c_1, c_2, c_3, \ldots, c_L], \quad \boldsymbol{C_i} \in \mathbb{R}^L \tag{6}$$

We use m kernels in the convolution operation, and there would be m feature maps, as is shown in the following, that represent the extracted local features:

$$[\boldsymbol{C_1}, \boldsymbol{C_2}, \boldsymbol{C_3}, \ldots, \boldsymbol{C_m}] \tag{7}$$

To avoid overfitting, SpatialDropout operation is employed in this paper. Dropout operation resets some elements as zeros randomly and independently, whereas SpatialDropout would randomly replace all values in a specific dimension with zeros, which can benefit the independence among feature maps. The distinction between Dropout and SpatialDropout is depicted in Fig. 2.

2) BIDIRECTIONAL GATED RECURRENT UNIT (BGRU)

RNN can process sequential data and learn long-term dependencies in sequential learning, while avoiding gradient vanish and gradient explosion during learning. Therefore, we take the feature maps obtained from convolution operations as the input of RNN. In this paper, we utilize the BGRU that consists of double layers of GRU[27], one of the layers learns features from the sequential data, and the other learns features from the reversed input data. GRU introduces the reset gate and the update gate to control its input and output.

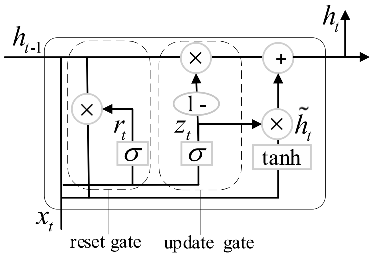

FIGURE 3. Structure of GRU.

See Fig. 3 for the structure of GRU, where the input at time t is x_t, h_{t-1}, the output is h_t, y_t, and the parameters are updated by the following formulas:

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right) \tag{8}$$

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right) \tag{9}$$

$$\tilde{h}_t = tanh\left(W_h \cdot \left[r_t \times h_{t-1}, x_t\right]\right) \tag{10}$$

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \tag{11}$$

$$y_t = \sigma\left(W_o \cdot h_t\right) \tag{12}$$

where $[\boldsymbol{x}, \boldsymbol{y}]$ is the concatenation of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$; r_t and z_t correspond to the reset gate and the update gate, respectively; $\tilde{h}_t$ denotes the candidate hidden layer; $\bullet$ denotes the element-wise multiplication; σ denotes the sigmoid function of which the formula is (13); $tanh$ denotes the $tanh$ function of which the formula is shown in (14); and W_r, W_z, W_h, W_o are the parameters of GRU.

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{13}$$

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{14}$$

The structure above enables the model to learn context information adequately from the input sequential data. And the dropout operation which randomly zeroes some of the neuron elements during each training process is an effective way for regularization.

3) POOLING LAYER, CONCATENATION LAYER AND FULLY CONNECTED LAYER WITH SOFTMAX OUTPUT

In this paper, we apply global-max-pooling and global-average-pooling operations over sequential data to transform feature maps of size L × G into size G, in which p_{ave}^i is the average of the i-th row and p_{max}^i is the max of the i-th row.

$$\boldsymbol{P_{ave}} = \left[\boldsymbol{p_{ave}^1}, \boldsymbol{p_{ave}^2}, \boldsymbol{p_{ave}^3}, \ldots, \boldsymbol{p_{ave}^G}\right], \quad \boldsymbol{P_{ave}} \in \mathbb{R}^G \tag{15}$$

$$\boldsymbol{P_{max}} = \left[\boldsymbol{p_{max}^1}, \boldsymbol{p_{max}^2}, \boldsymbol{p_{max}^3}, \ldots, \boldsymbol{p_{max}^G}\right], \quad \boldsymbol{P_{max}} \in \mathbb{R}^G \tag{16}$$

Then, concatenate the results of pooling operations:

$$\boldsymbol{Z} = \boldsymbol{P_{ave}} \oplus \boldsymbol{P_{max}}, \quad \boldsymbol{Z} \in \mathbb{R}^{2 \times G} \tag{17}$$

The calculated feature map $\boldsymbol{Z}$ is then passed to a fully connected layer followed by a softmax output function, and the final output is the probability distribution over all categories. The softmax function is as follows:

$$Q_i(o) = \frac{e^{o_i}}{\sum_{i=1}^{m} e^{o_i}} \tag{18}$$

In this work, we choose cross entropy as the loss function to measure the divergence between actual values and predictions of the model and to reduce losses through iterations.

IV. EXPERIMENTS

All experiments reported in this section are conducted on GPUs using Keras (https://keras.io). And the Python 3.6 programming language under Ubuntu system. For each model, we use categorical cross-entropy as the loss function and Adam optimizer with $lr = 0.001$, $\beta_1 = 0.9, \beta_2 = 0.999$ and $\epsilon = 1e - 8$.

A. EXPERIMENT DATASET

1) DATA DESCRIPTION

We conduct experiments on datasets from AI Challenger 2018 (*https://ai.chuangxin.com/ai_challenger*) fine-grained user review sentiment analysis competition (Corp1) and DataFountain (*https://www.datafountain.cn/*) user opinion, topic and sentiment recognition competition of the automotive industry (Corp2).

In Corp1, there are sentiments with 20 fine-grained aspects in 6 categories. Reviews in the dataset are graded into two layers deferring from their granularities. The first layer contains coarse-grained entities and the second layer contains fine-grained aspects. A concrete grade assignment is shown in Table 3. Four values $[1, 0, -1, -2]$ are used to demonstrate the fine-grained sentiment polarities of positive, neutral, negative and unmentioned.

In Corp2, reviews involve 10 attributes of cars and the number of reviews on each attribute is depicted in Table 1. In this paper, we take the most commonly used attributes (power, price, fuel consumption, maneuverability, and comfort) to testify the model. The dataset is labeled the same as Corp1.

Preprocess review texts in these two datasets with the word segment tool Jieba for tokenization and then remove the stop words in them. Details of the two datasets are shown in Table 2.

2) DATA LABELING

For the Corp1 dataset, in this work, we mainly measure the attributes or aspects in its first layer. Since the original datasets only gives the sentiment polarity towards each sub-attribute in the second layer, the data labels should be processed to obtain the label of reviews related to each attribute in the first layer. The reviews with each attribute are labeled as:

$$\left\{\begin{array}{l} -2\ if\ all \begin{pmatrix} labels\ in\ the\ second\ layer \\ under\ the\ first\ layer \end{pmatrix} = -2 \\ \left\{\begin{array}{l} 0\ if\ sum \begin{pmatrix} labels\ in\ the\ second\ layer \\ under\ the\ first\ layer \end{pmatrix} = 0 \\ 1\ if\ sum \begin{pmatrix} labels\ in\ the\ second\ layer \\ under\ the\ first\ layer \end{pmatrix} > 0 \\ -1\ if\ sum \begin{pmatrix} labels\ in\ the\ second\ layer \\ under\ the\ seconde\ layer \end{pmatrix} < 0 \end{array}\right\} \end{array}\right\} \tag{19}$$

TABLE 1. Number of labeled reviews of each aspect.

Attribute	Power	Price	Fuel consumption	Manipulation	Comfort
Number	2732	1273	1082	1036	931
Attribute	Configuration	Safety	Trim	Appearance	Space
Number	853	573	536	489	442

TABLE 2. Review information of experiment datasets.

	Hotel (Corp1)		Car (Corp2)	
	Char number in reviews	Word number in reviews	Char number in reviews	Word number in reviews
max	1779	1110	181	114
min	10	6	6	3
avg	292.116783	176.85975	39.3347	23.1175

TABLE 3. Aspect classification and labels of the given example.

The first layer	The second layer	Label of the first layer	Label of the second layer
Location	Traffic convenience	-2	-2
	Distance from business district	-2	
	Easy to find	-2	
Service	Wait time	-2	-2
	Waiter's attitude	-2	
	Parking convenience	-2	
	Serving speed	-2	
Price	Price level	-2	1
	Cost-effective	1	
	Discount	-2	
Environment	Decoration	1	1
	Noise	-2	
	Space	-2	
	Cleanness	1	
Dish	Portion	1	1
	Taste	1	
	Look	-2	
	Recommendation	-2	

As an example, Table 3 shows the processed labels of the review 'The noodle here tastes good and the restaurant is very cost-effective. The food portion is large. Girls who have a small appetite may not finish eating. The atmosphere is good among noodle restaurants. At least it looks bright and clean. Generally, other small shops' sanitary conditions are inferior to its. At lunchtime, even the sidewalks are crowded with people. The owner of the restaurant next door is said to be a relative with this noodle restaurant, and sometimes the restaurant is also open for people to eat noodles.'

In addition, the eventual data distribution of labeled aspects in the first layer and the sentiment towards each aspect is shown in Tables 4-5.

TABLE 4. Experiment dataset of Corp1.

		Location	Service	Price	Environment	Dish
Unmentioned		67684	36266	35485	42561	3715
Mentioned	Positive	44678	52331	43700	55847	78519
	Negative	4143	14430	13720	9965	12057
	Neutral	3495	16973	27095	11627	25709
Total		120000	120000	120000	120000	120000

TABLE 5. Experiment dataset of Corp2.

		Power	Price	Fuel consumption	Manipulation	comfort
Unmentioned		5558	7017	7208	7254	7359
Mentioned	Neutral	1970	1014	793	606	564
	Positive	384	114	151	306	111
	Negative	378	145	138	124	256
Total		8290	8290	8290	8290	8290

B. EVALUATION METRICS

We report the accuracy, F1-score and area under the curve (AUC) of the proposed model, which are widely used as the performance evaluation metrics in multilabel classification tasks. Given a dataset D, set the length of D as N, and the number of entities that are correctly predicted as category i, actually in category i but miss-predicted as other categories, correctly predicted as another category and actually in another category but miss-predicted as category i are set as TP_i, FN_i, TN_i and N_i, respectively. Additionally, the calculation of accuracy and F1-score is shown as follows:

$$Acc = \frac{TP_i}{TP_i + FN_i + FP_i} \tag{20}$$

$$Weighted\_P = \frac{1}{N}\sum\nolimits_{i=1}^{n} \frac{TP_i}{TP_i + FP_i} \times N_i \tag{21}$$

$$Weighted\_R = \frac{1}{N}\sum\nolimits_{i=1}^{n} \frac{TP_i}{TP_i + FN_i} \times N_i \tag{22}$$

$$Weighted\_F1 = 2 \times \frac{Weighted\_P \times Weighted\_R}{Weighted\_P + Weighted\_R} \tag{23}$$

where $Weighted\_P$ denotes the weighted precision and $Weighted\_R$ denotes the weighted recall. Draw the Receiver Operation Characteristic (ROC) curve with False positive rate (FPR) and True positive rate (TPR). Here, the x-axis is FPR, which represents the probability of misclassifying an entity into category i, and the y-axis is TPR, which represents the probability of correctly classifying an entity into category i. The size of the area under the ROC curve is the Area under curve (AUC) value. In each category, we can calculate the FPR and TPR value at a certain threshold, then the ROC

TABLE 6. Experiment result of the proposed model.

		Acc	F1	AUC
Aspect	location	85.21%	85.08%	84.31%
	service	88.98%	89.11%	88.32%
	price	88.43%	88.62%	87.74%
	environment	88.08%	88.11%	87.23%
	dish	97.33%	96.98%	67.70%
Sentiment	location	87.82%	85.77%	90.87%
	service	74.98%	73.83%	81.24%
	price	71.78%	71.76%	78.84%
	environment	79.03%	78.53%	84.27%
	dish	74.17%	71.28%	80.63%

curves are drawn. In this way, a total of n AUC values is obtained, and we take these AUC values to obtain the final AUC value.

C. HYPERPARAMETERS

We use word embeddings from word2vec as the input of the model, initialize related hyperparameters and set the dimension of the word embeddings as 100. The number of kernels in the convolution neural network is 3 in 64 filters, and the units of the GRU is 32. Divide the dataset at a proportion of 8:1:1. The experimental results of the proposed model are shown in Table 6.

To verify the differences of the aspects of hotels, we conduct a significant test on the experiment result in Table 6 with the Friedman test method [28]. Assuming that there are no significant differences among the five aspects in Table 6, by calculation, we got the $\rho = 0.024 < 0.05$, which proves that at a significance of $\alpha = 0.05$, the five aspects of hotels are of significant differences.

To better understand the effect of hyperparameters on model performance, we conduct extensive experiments on Corp1 to examine the influences of the word-embedding dimension, filter size, GRU units on the experiment's performance with three evaluation metrics, and the results are shown in Fig. 4-Fig. 7.

Fig. 4 depicts the performances of the proposed model with different word embedding dimensions over three evaluation metrics. It shows that when the dimension is set too small, the word representation may contain too little information, which can lead to underfitting. Whereas, when the dimension is too large, the number of parameters in the CNN model will increase rapidly and the explosion of parameters will raise the risk of overfitting and decelerate the training process.

Fig. 5 and Fig. 6 describe the performances of the proposed model with different numbers of filters and filter sizes in the convolutional layer. During the convolution operation, the

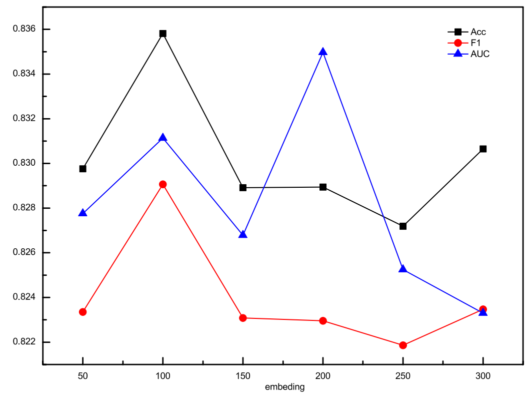

FIGURE 4. Effect of the word embedding dimension.

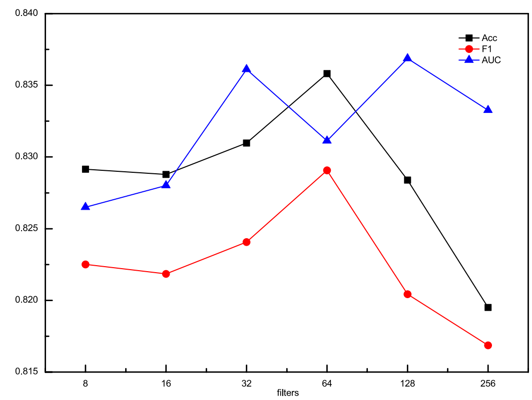

FIGURE 5. Effect of the filter number.

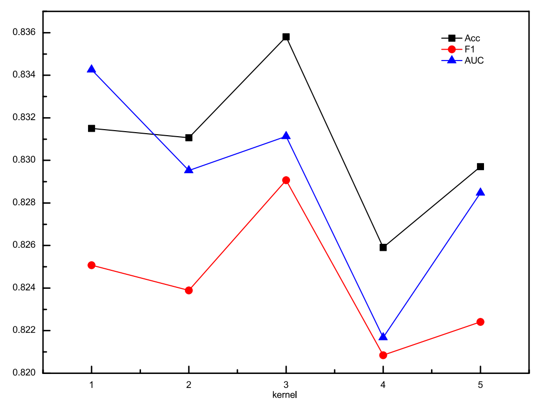

FIGURE 6. Effect of the filter size.

proposed model automatically learns N-gram features from data by deep learning and then input the learned information into BGRU for further analysis. With the increase in the filter size, the extracted local features are enriched gradually. However, excessively rich information can bring noises that will affect the prediction accuracy of the model.

Fig. 7 shows the effect of GRU units on the proposed model. BGRU can learn features from both sequential and reversed input data. If the output units are too small, the information in feature maps cannot be fully utilized, whereas, if the output units are too large, the overfitting risk of the

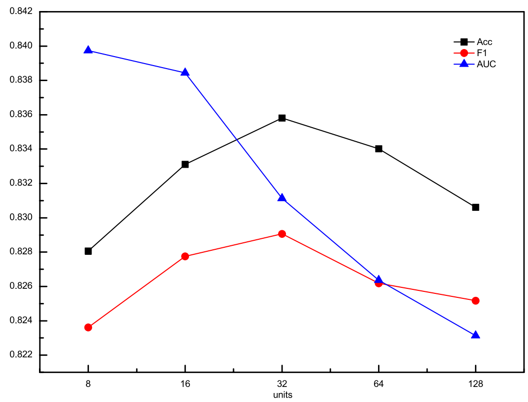

FIGURE 7. Effect of the GRU units.

TABLE 7. Comparative experiment result.

	Aspect			Sentiment		
	Acc	F1	AUC	Acc	F1	AUC
our model	89.61 %	89.58 %	83.06 %	77.56 %	76.23 %	83.17 %
TextCNN	89.12 %	89.04 %	81.68 %	76.35 %	75.94 %	82.26 %
DE_CNN	86.78 %	86.94 %	81.22 %	71.94 %	71.74 %	78.96 %
Dis_LSTM	87.85 %	87.80 %	80.24 %	71.36 %	70.51 %	78.52 %
CNN_LSTM	88.15 %	87.88 %	78.32 %	70.46 %	72.08 %	77.85 %
SVM	87.63 %	87.44 %	78.37 %	73.20 %	65.68 %	79.90 %

model is increased and the training time needed will be extended.

D. COMPARATIVE EXPERIMENT

To verify the effectiveness of the proposed model in this paper, we compare it with several existing models, including TextCNN [10], DE_CNN [29], Dis_LSTM [30], CNN_LSTM [14], and SVM [31] on Corp1. The results of the comparisons are shown in Table 7, and from them, we can notice that our model outperforms the other models across all three metrics, which validates the effectiveness of the proposed model.

Assuming that there are no significant differences between the models in Table 7, we perform significant test with Friedman test method [28] on the models. After calculation, we got a $\rho = 0.024 < 0.05$, which proves that at a significance of $\alpha = 0.05$, the models are of significant differences.

E. DOMAIN EXPANSION CAPABILITY

We test the proposed model on Corp2 with the parameters set the same as the experiments on Corp1 to further verify its domain expansion capability, and the experiment results are shown in Table 8. As seen from Table 8, the model in this paper achieves a good performance on fine-grained sentiment analysis in the automotive field. Specifically, the model does better on aspect recognition in the automotive field than in the hotel field, whereas, its sentiment polarity identification performance in the automotive field is slightly worse than that in the hotel field. Overall, the model shows a comparable capability in both of the fields, which verifies its domain expansion capability.

TABLE 8. Domain expansion capability experimental result.

	Aspect			Sentiment		
	Acc	F1	AUC	F1	Acc	AUC
hotel	89.61 %	89.58 %	83.06 %	77.56 %	76.23 %	83.17 %
car	93.82 %	93.64 %	86.93 %	69.89 %	64.02 %	77.42 %

V. CONCLUSION

Research on online reviews has attracted considerable attention, and ABSA has always been one of the important tasks in natural language processing. Therefore, it is highly necessary to explore different sentiment polarities in reviews. For this reason, we propose a deep neural network model for aspect-based sentiment analysis utilizing the combination of CNN and GRU. The characteristics of the model are as follows: obtaining local features in reviews and sequential relations through CNN; learning long-tern dependency and location relation by RNN; and extracting global features from whole sentences with global-max-pooling and global-average-pooling. The extensive experiments demonstrate that the proposed model performs well on hotel and car datasets and achieves a high accuracy for aspect extraction and sentiment classification. The AUC values of these two subtasks reach 83.06% and 83.17% on the hotel dataset and 86.93% and 77.42% on the car dataset. This paper mainly focuses on analyzing sentiments of multiple aspects towards a certain entity. In further study, we will do more detailed investigation and explore the situation in which several entities are mentioned in one review. Besides, some of the most representative computational intelligence algorithms can also be used on ABSA, such as Monarch Butterfly Optimization (MBO) [32]–[35], Earthworm Optimization Algorithm (EWA) [36], Elephant Herding Optimization (EHO) [37]–[39] and Moth Search (MS) [40], [41] algorithm. Thus, in our future study, we will look into the use of theses algorithms on ABSA research.

REFERENCES

[1] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, *A Practical Guide to Sentiment Analysis*. Cham, Switzerland: Springer, 2017.

[2] B. Liu, "Sentiment analysis and opinion mining," *Synth. Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.

[3] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima, "Mining product reputations on the Web," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining - KDD*, 2002, pp. 341–349.

[4] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, and V. Hoste, "SemEval-2016 task 5: Aspect based sentiment analysis," in *Proc. 10th Int. Workshop Semantic Eval., SemEval*, San Diego, CA, USA: Association for Computational Linguistics, Jun. 2016, pp. 19–30.

[5] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, 2016, pp. 260–270.

[6] A. M. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization," 2015, *arXiv:1509.00685*. [Online]. Available: https://arxiv.org/abs/1509.00685

[7] X. Y. Wang, W. J. Jiang, and Z. Y. Luo, "Combination of convolutional and recurrent neural network for sentiment analysis of short texts," in *Proc. COLING, Tech. Papers*, Dec. 2016, pp. 2428–2437.

[8] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning, "Semi-supervised recursive autoencoders for predicting sentiment distributions," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Edinburgh, U.K., 2011, pp. 151–161.

[9] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1422–1432.

[10] D. Santos, C. Nogueira, and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proc. COLING, Tech. Papers*, 2014, pp. 69–78.

[11] W. Yu and W. N. Zhou, "Sentiment analysis of commodity reviews based on LSTM," *Comput. Sci. Appl.*, vol. 27, no. 08, pp. 159–163, 2018.

[12] O. Araque, I. Corcuera-Platas, J. F. Sánchez-Rada, and C. A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," *Expert Syst. Appl.*, vol. 77, pp. 236–246, Jul. 2017.

[13] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 3111–3119.

[14] Y. Kim, "Convolutional neural networks for sentence classification," 2014, *arXiv:1408.5882*. [Online]. Available: http://arxiv.org/abs/1408.5882

[15] X. Zhu, P. Sobhani, and H. Guo, "Long short-term memory over recursive structures," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 2, 2015, pp. 1604–1612.

[16] Y. Ren, R. Wang, and D. Ji, "A topic-enhanced word embedding for Twitter sentiment classification," *Inf. Sci.*, vol. 369, pp. 188–198, Nov. 2016.

[17] I. Chaturvedi, E. Ragusa, P. Gastaldo, R. Zunino, and E. Cambria, "Bayesian network based extreme learning machine for subjectivity detection," *J. Franklin Inst.*, vol. 355, no. 4, pp. 1780–1797, Mar. 2018.

[18] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, 2014, pp. 27–35.

[19] V. Perez-Rosas, C. Banea, and R. Mihalcea, "Learning sentiment lexicons in Spanish," in *Proc. 8th Int. Conf. Lang. Resour. Eval. (LREC)*, 2012, pp. 3077–3081.

[20] S. M. Mohammad, S. Kiritchenko, and X. Zhu, "NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets," 2013, *arXiv:1308.6242*. [Online]. Available: https://arxiv.org/abs/1308.6242

[21] Y. Meng, H. W. Wang, and W. Wang, "The effect of electronic word-of-mouth on sales through fine-gained sentiment," *Manage. Rev.*, vol. 29, no. 1, pp. 144–154, 2017.

[22] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014.

[23] W. Xue, W. Zhou, T. Li, and Q. Wang, "MTNA: A neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews," in *Proc. 8th Int. Joint Conf. Natural Lang. Process.*, no. 2, 2017, pp. 151–156.

[24] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "Adaptive recursive neural network for target-dependent Twitter sentiment classification," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics (Short Papers)*, vol. 2, 2014, pp. 49–54.

[25] B. Liang, L. Quan, X. Jin, Z. Qian, and Z. Peng, "Aspect-based sentiment analysis based on multi-attention CNN," *J. Comput. Res. Develop.*, vol. 54, no. 8, p. 1724, 2017.

[26] Y. Xiao and G. Zhou, "Syntactic edge-enhanced graph convolutional networks for aspect-level sentiment classification with interactive attention," *IEEE Access*, vol. 8, pp. 157068–157080, 2020.

[27] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*. [Online]. Available: http://arxiv.org/abs/1412.3555

[28] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *J. Amer. Stat. Assoc.*, vol. 32, no. 200, pp. 675–701, Dec. 1937.

[29] H. Xu, B. Liu, L. Shu, and P. S. Yu, "Double embeddings and CNN-based sequence labeling for aspect extraction," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, no. 2, 2018, pp. 592–598.

[30] D. Yogatama, C. Dyer, W. Ling, and P. Blunsom, "Generative and discriminative text classification with recurrent neural networks," 2017, *arXiv:1703.01898*. [Online]. Available: http://arxiv.org/abs/1703.01898

[31] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proc. Conf. Empirical Methods Natural Language Process.*, 2002, pp. 79–86.

[32] G. G. Wang, S. Deb, and Z. Cui, "Monarch butterfly optimization," *Neural Comput. Appl.*, vol. 31, no. 7, pp. 1995–2014, 2019.

[33] Y. Feng, S. Deb, G.-G. Wang, and A. H. Alavi, "Monarch butterfly optimization: A comprehensive review," *Expert Syst. Appl.*, vol. 168, Apr. 2021, Art. no. 114418.

[34] Feng, Yu, and Wang, "A novel monarch butterfly optimization with global position updating operator for large-scale 0-1 knapsack problems," *Mathematics*, vol. 7, no. 11, p. 1056, Nov. 2019.

[35] Y. Feng, G.-G. Wang, W. Li, and N. Li, "Multi-strategy monarch butterfly optimization algorithm for discounted 0-1 knapsack problem," *Neural Comput. Appl.*, vol. 30, no. 10, pp. 3019–3036, Nov. 2018.

[36] G.-G. Wang, S. Deb, and L. dos S. Coelho, "Earthworm optimisation algorithm: A bio-inspired metaheuristic algorithm for global optimisation problems," *Int. J. Bio-Inspired Comput.*, vol. 12, no. 1, pp. 1–22, Jan. 2018.

[37] G.-G. Wang, S. Deb, and L. D. S. Coelho, "Elephant herding optimization," in *Proc. 3rd Int. Symp. Comput. Bus. Intell. (ISCBI)*, Dec. 2015, pp. 1–5.

[38] G. G. Wang, S. Deb, X.-Z. Gao, and L. D. S. Coelho, "A new metaheuristic optimisation algorithm motivated by elephant herding behaviour," *Int. J. Bio-Inspired Comput.*, vol. 8, no. 6, pp. 394–409, 2016.

[39] J. Li, H. Lei, A. H. Alavi, and G.-G. Wang, "Elephant herding optimization: Variants, hybrids, and applications," *Mathematics*, vol. 8, no. 9, p. 1415, Aug. 2020.

[40] G.-G. Wang, "Moth search algorithm: A bio-inspired metaheuristic algorithm for global optimization problems," *Memetic Comput.*, vol. 10, no. 2, pp. 151–164, Jun. 2018.

[41] Y.-H. Feng and G.-G. Wang, "Binary moth search algorithm for discounted 0-1 knapsack problem," *IEEE Access*, vol. 6, pp. 10708–10719, 2018.

NARISA ZHAO received the B.S. degree in applied mathematics, the M.S. degree in professional operations research and cybernetics, and the Ph.D. degree in system engineering from the Dalian University of Technology, in 1992, 1995, and 1998, respectively.

From April 1998 to April 1999, he worked with the Construction Land Examination and Approval Office of Dalian Urban and Rural Planning and Land Bureau, as a Chief Staff Member. Since April 1999, he has been involved in teaching and scientific research with the Department of Management and Economics, Dalian University of Technology, China, where he also served as a Lecturer, an Associate Professor, and a Professor. He is currently a Ph.D. Supervisor with the Institute of Systems Engineering, Dalian University of Technology. His research interests include social networks, text mining and sentiment analysis, and electronic commerce.

HUAN GAO was born in Mudanjiang, Heilongjiang, China, in 1994. She received the B.S. degree in management from the University of Science and Technology Beijing, China, in 2017, and the M.S. degree in management science and engineering from the Dalian University of Technology, China, in 2020.

Her research interest includes text mining. She is currently a Peer Reviewer of *Applied Intelligence* (APIN).

XIN WEN was born in Hohhot, China, in 1996. She received the B.S. degree in management from the Dalian University of Technology of China, in 2019, where she is currently pursuing the M.S. degree in management science and engineering.

Her research interest includes online opinion mining.

HUI LI was born in Dalian, China, in 1988. She received the B.S. and M.S. degrees from Dalian Maritime University, in 2010 and 2012, respectively, both in transportation engineering, and the Ph.D. degree in management science from the Dalian University of Technology, in 2020.

From 2018 to 2020, she was sponsored as a Visiting Scholar in marketing with the University of Alberta by the China Scholarship Council. Her research interests include data mining in social media, online WOM sentiment analysis, and information diffusion on social networks.