IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 2, 2018, accepted December 18, 2018, date of publication January 1, 2019, date of current version January 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2890390

Multiple Level Hierarchical Network-Based Clause Selection for Emotion Cause Extraction

XINYI YU[1,2], WENGE RONG[1,3], (Senior Member, IEEE), ZHUO ZHANG[1,2], YUANXIN OUYANG[1,3], AND ZHANG XIONG[1,3]

[1]State Key Laboratory of Software Development Environment, Beihang University, Beijing 100191, China
[2]Shenyuan Honoary School, Beihang University, Beijing 100191, China
[3]School of Computer Science and Engineering, Beihang University, Beijing 100191, China

Corresponding author: Wenge Rong (w.rong@buaa.edu.cn)

This work was supported in part by the State Key Laboratory of Software Development Environment of China under Grant SKLSDE-2017ZX-16, and in part by the National Natural Science Foundation of China under Grant 61332018.

ABSTRACT Emotion cause extraction is one of the most important applications in natural language processing tasks. It is a difficult challenge due to the complex semantic information between emotion description and the whole document. Previous approaches have revealed that clause is an important indicator of emotion-cause extraction. As such, selecting a suitable clause has become an interesting challenge. Different from existed clause selection methods which mainly focus on semantic similarity between clause and emotion description, in this paper, we proposed a hierarchical network-based clause selection framework in which the similarity is calculated by considering document features from word's position, different semantic levels (word and phrase), and interaction among clauses, respectively. Experimental study on a Chinese emotion-cause corpus has shown the proposed framework's effectiveness and the potential of integrating different level's information.

INDEX TERMS Emotion cause extraction, hierarchical network, clause selection, attention.

I. INTRODUCTION

Emotion cause extraction is a fundamental and challenging task for emotion analysis and deserves well-investigated [1]. Different from other emotion tasks such as emotion classification [2], emotion recognition [3] and reader's emotion prediction [4], emotion-cause extraction not only focuses on emotion expression, but also cares about the stimuli of an emotion [1]. The cause of emotion is sometimes more important than the emotion itself in certain scenarios. For example, in the social network based recommendation websites, there are a large number of evaluation and feedback from users. The service provider, e.g, restaurant, cares more about why costumers like or dislike their service rather than emotion included in the comments.

Figure 1 is an emotion extraction example from SINA City News, where ''sad'' is an emotion word and the cause of ''sad'' is ''she was told the bad news about the death of her husband''. The goal of emotion-cause extraction task is to identify the reason behind an emotion expression [5]. Extracting emotion cause from texts requires better comprehension of the text and more features learned from the sentences.

邮递员给李太太带来了前线的信件，
The postman brought a letter to Mrs. Li.
当她迫不及待地拆开邮件时，
When she opened the mail,
却被告知丈夫殉职的噩耗。
she was told the bad news about the death of her husband.
李太太顿时崩溃大哭，伤心地晕了过去
Mrs. Li suddenly burst into tears and became too sad to faint.

FIGURE 1. Emotion document example (Each sentence's English translation is listed below the Chinese sentence).

In the earlier study, some researchers tried to develop a set of rules to help identify the emotion cause. For example, Li and Xu [6] first constructed an automatic rule-based system to detect the event cause of each document. Later on, Gao *et al.* [7] also adopted a rule-based approach to detect emotion cause. Their method extracted various linguistic cues from an emotion cause annotated corpus and then generalized a list of linguistic rules with the help of these cues. Unlike conventional statistical based methods,

their approach inferred and extracted the reasons of emotions by importing knowledge and theories from other fields, e.g., sociology area. Though rule based approaches are easy in implementation, their performance heavily relies on the size of annotated corpus and these approaches are less flexible.

Some researchers started to analyze the syntactic information of the sentences in the document to help extract emotion cause. For example, Gui *et al.* [8] proposed a syntactical tree based multi-kernel SVMs model to complete this task. Their method could take lexical features into consideration and also detect all possible combinations of syntactic structures to obtain sufficient features for emotion analysis with the use of a limited training set. This method achieved better performance compared to the previous study, while it would be more beneficial to consider the semantic information between emotion description and rest sentences [9].

As to the semantic perspective, Gui et al. revealed that the presence of conjunctions and prepositions could indicate the discourse information among clauses, which suggests the importance of clause level information. The efficiency of such clause level information has been also proven in other emotion analysis fields. For example, the experimental results conducted in emotion annotation have shown clause-level features can improve the prediction of emotion of the sentence [10].

Therefore, the basic analysis unit should also consider clause level information [8]. Gui *et al.* [5] argued that it is an initial and important step to split the context into several clauses and each clause maybe considered as a text containing the description of an event which may or may not cause the certain emotion. Based on this idea, they proposed a Convolutional Multiple-Slot Memory Network (ConvMS-Memnet) to deal with the clause selection by considering the relationship between each clause and the emotion description. This model could extract both word level sequence features and lexical features by storing relevant context in different memory slots. Although this method has achieved great performance by taking the relationship between clause and emotion description into consideration, it might be more useful for emotion extraction if we can also consider the relationship among clauses because just modeling the semantic matching between individual pair of emotion description and a single clause maybe have limited influence [11].

In light of these challenges, we proposed a multiple level hierarchical network based clause selection (HCS) approach. This hierarchical framework is inspired by the hierarchical model in another type of sentiment analysis task [12], which implemented word-level and clause-level attentions for aspect sentiment classification and highlighted the efficiency of incorporating the importance degrees of both words and clauses inside a sentence. In this research, we propose to take word's position, different levels of semantic information (i.e., word-level and phrase-level), as well as relationship among clauses all into consideration for emotion-cause extraction. The goal is to more effectively find the clause which obtains the highest probability to be the answer of the emotion. In this model, a word level attention network, which consists of content attention and position attention [13], is first implemented to model the word-level information. Afterwards convolutionary neural network (CNN) is utilized for analyzing phrase level information with regard to emotion description. According to the Statistical MT (SMT), phrases are simply chains of words that frequently co-occur and are aligned with the same source word sequences [14]. Besides, the bidirectional gated recurrent units (Bi-GRU) is adopted to extract relationships among clauses to help clause selection. The experimental results have shown potential of this hierarchical framework in clause selection.

The rest of the paper is organized as follows. Section 2 proposed a multiple level hierarchical network based model for emotion-cause extraction (HCS). Section 3 will study the experimental results and Section 4 will discuss the related work in this area. Finally, Section 5 concludes the work and outlines the future directions.

II. METHODOLOGY

Figure 2 is the proposed hierarchical framework in which the emphasis is put on different level of the emotion document gradually during the whole clause selection process. First, an attention based word level network will be conducted to study words' content information and position's influence in the emotion document. Afterwards a CNN based phrase level network is proposed to further investigate the relation between phrase and emotion description. Finally the interaction among clauses will be studied by a Bi-GRU based clause level network.

A. DATA REPRESENTATION

In a clause selection task, a given emotion document D contains an emotion description E and other clauses [5]. For each document $D = \{c_1, \ldots c_i, \ldots, c_n\}$ consisting of n clauses, the goal is to identify which clause contains the emotion-cause. For an input clause $c_i = \{w_1^i, \ldots, w_j^i, \ldots, w_m^i\}$, where each word w_j^i will be mapped into a dimensional embedding [15]. The word embedding matrix can be represented as $Emb \in R^{d \times |V|}$, where d is the dimension of word vector and V is the vocabulary size. Similarly, the words in emotion description E are also represented by word embedding. In the implementation of the model, it will be convenient to pad the emotion description and all of the clauses included in the emotion document to have the same length $s = max(c_1, c_2, \ldots, c_n, E)$. As a result, each clause including emotion description can be represented as a feature map of dimension $d \times s$.

B. WORD LEVEL ATTENTION NETWORK

The attention mechanism in this part consist of content attention and position attention. The basic intuition of content attention is that words in each clause does not contribute equally to the semantic meaning of a clause, and the position attention is designed for the reason that intuitively a context

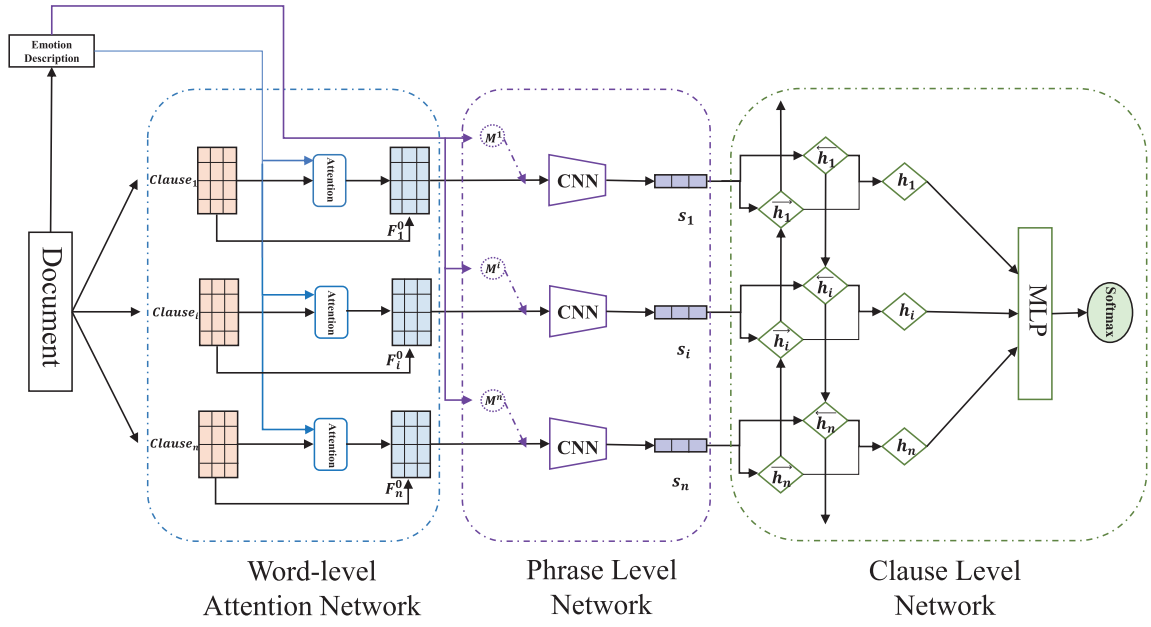

FIGURE 2. Multiple level hierarchical network based clause selection framework.

word closer to the emotion description should be attached more importance than a farther one [16].

1) CONTENT ATTENTION

This kind of attention is proposed to compute an alignment score between elements from two sources [17]. As show in Fig. 2 , each row of a $clasue_i$ is the representation of a word. We describe the attention feature matrix between emotion description E and the i^{th} clause c_i as A_i whose attention weights are computed on the output of convolution layer. Let $V_i = \{v_1^i, v_2^i, \ldots, v_m^i\}$ and $(v_j^i \in R^d)$ denote the vectors of the words in a clause, in this part, attention computes the alignment score between v_j^i and E according to a compatibility function $f(v_j^i, E)$, which represents the attention of E to v_j^i. Then a softmax function is implemented to compute the probability distribution $p(z|V_i, E)$, where z indicates the importance of the token in clause c_i to emotion description E. The process above can be summarized by equations as follows:

$$a_j^i = [f(v_j^i, E)]_{j=1}^n \tag{1}$$

$$p(z_i = j|V_i, E) = \frac{\exp(f(v_j^i, E))}{\sum_{j=1}^{m} \exp(f(v_j^i, E))} \tag{2}$$

where $f(v_j^i, E)$ can be represented as follows:

$$f(v_j^i, E) = W^T \sigma(W^{(1)} v_j^i + W^{(2)} E) \tag{3}$$

The $\sigma(\cdot)$ is an activation function [17] and W^T is a weight vector. The $W^{(1)}$ and $W^{(2)}$ are weighted vectors. Then the output of clause c_i after attention layer can be represented as:

$$F_i^0 = V_i \cdot A^i \tag{4}$$

where $F_i^0 \in R^{m \times d}$.

2) POSITION ATTENTION

The content attention mechanism mentioned above ignores the position information between word in clauses and emotion description. In this part, three strategies are adopted to encode the location information in the attention model. The details are described as follows:

- Strategy 1. Following the position encoding method proposed in [13], we define the position sequence relative to the emotion description E, where

$$p_j^i = \begin{cases} \dfrac{i - e_1}{n_i} & i < e_1 \\ 0 & e_1 < i < e_2 \\ \dfrac{e_2 - 1}{n_i} & i > e_2 \end{cases}$$

  Here $p_j^i \in R^{1 \times m}$ is the relative distance of token v_j^i in clause c_i to the emotion description E; n_i is the clause length; and e_1, e_2 are the starting and ending indices of the emotion description E respectively. Then we pad vector p_j^i into a matrix $P^i \in R^{m \times m}$ and add position attention to the previous model with:

$$F_i^1 = P^i \cdot F_i^0 \tag{5}$$

  where $\cdot$ means element-wise multiplication and $F_i^1 \in m \times d$.
- Strategy 2. Compared to the attention calculation conducted in Strategy 1, we adopt another method to represent the position sequence p_j^i as

$$p_j^i = 1 - \frac{l_i}{n} \tag{6}$$

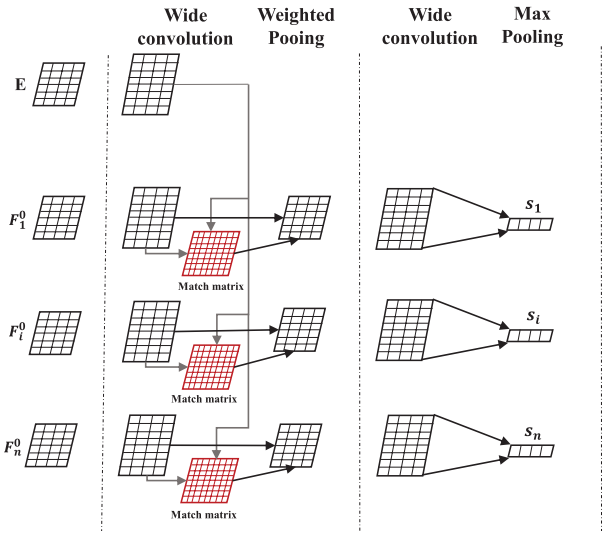

FIGURE 3. Weighted pooling layer based CNN.

where n is the document length, l_i is the location of word w_i in the whole document.

- Strategy 3. Inspired by the location model used in [16], we consider the location vector p_j^i of word w_j^i in clause c_i as a parameter and regard location representations as neural gates. These gates are able to control how many semantics should be stored for the next stage. The clause representation after this strategy can be shown as

$$F_i^1 = \sigma(P^i) \cdot F_i^0 \tag{7}$$

where position vectors are fed into a sigmoid function σ.

C. PHRASE LEVEL NETWORK

Though word is an important hint in emotion-cause detection, the phrase is also an import indicator for detecting the relationship between the clause and the emotion description. As such in this research we further propose a CNN based phrase level relation detection part, as shown in Fig. 3.

1) CONVOLUTIONAL LAYER

The convolutional filter can be represented as $W_i^l \in R^{l \times d}$, where l is the window size of the filter. Let $F_i^0 = \{f_{i,1}^0, \ldots, f_{i,j}^0, \ldots, f_{i,m}^0\}$, then a representation for each phrase of length l can be learned.

$$P_{i,j}^l = Relu(W_i^l[f_{i,j}^0, f_{i,j-1}^0, \ldots, f_{i,j-l+1}^0] + b_i^l) \tag{8}$$

here we use $P_{i,j}^l$ to represent phrase vectors generated by convolutional layer, which includes all phrases ended with $f_{i,j}^0$. The number of words in phrase is equal to the size of filter.

2) WEIGHTED POOLING LAYER

We implement a match matrix which compares all units in c_i with all units of E in order to re-weight the convolution output. The match matrix between clause c_i and E can be represented as follow:

$$M_{j,k}^i = f(F_i^0[:, j], E[:, k]) \tag{9}$$

here the function f can be defined as: $f(x, y) = 1/(1+|x-y|)$, where $|\cdot|$ is Euclidean distance [18].

The scores of column j in M^i stand for the attention distribution of the j-th unit in E with respect to c_i, and the scores of row k in M^i denote the attention distribution of the k-th unit in c_i with respect to E. Let $m_j^i = \sum M^i[j, :]$ be the weight of unit j in c_i and $P_i \in R^{(s+w-1)\times d}$ be the output of convolutional layer for clause c_i. Then the new feature map after weighted pooling layer can be computed by the following equation:

$$F_i^p[j, :] = \sum_{k=j:j+w} m_k^i P_i[k, :], j = 1, \ldots, s \tag{10}$$

where $F_i^p \in R^{s \times d}$.

D. CLAUSE LEVEL NETWORK

We denote the representation of clause c_i after convolutional neural network as s_i ($s_i \in R^{1 \times d_1}$). To capture the information among the clauses, the bidirectional gated recurrent units (Bi-GRU) are used. Bi-GRU can capture the sematic correlations among the clauses which are from the same document, as GRU is easier to capture long-term dependencies and has more persistent memory compared to RNN. GRU is considered as a simplified LSTMs, which based on just two multiplicative gates [19]. The GRU model can be defined by the following equations:

$$z_t = \sigma W_z x_t + U_z h_{t-1} + b_z \tag{11}$$

$$r_t = \sigma W_r x_t + U_r h_{t-1} + b_r \tag{12}$$

$$\tilde{h_t} = tanh(\sigma W_h x_t + U_h(h_{t-1} \odot r_t) + b_h) \tag{13}$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h_t} \tag{14}$$

The network is fed with input vector x_t with parameters $\{W_z, W_r, W_h, U_z, U_r, U_h\}$ and bias $\{b_z, b_r, b_h\}$. z_t, r_t, h_t represent the update gate, reset gate, hidden layer respectively and $\tilde{h_t}$ is the candidate state obtained with a hyperbolic tangent [20]. The activations of both update gate and reset gate are element-wise logistic sigmoid functions $\sigma(\cdot)$, $\odot$ stands for element-wise multiplication that can be represented as $\sigma(x) = \frac{1}{1+e^{-x}}$, and $tanh(\cdot)$ is a tangent function which can be defined as $tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.

The Bi-GRU model implemented in our experiment is based on the GRU model, which maintains two hidden layers, one for the left-to-right propagation, and another for the right-to-left propagation [21]. We use the sentence representation $S = \{s_1, \ldots, s_i, \ldots s_n\}$ obtained from CNN model as the inputs of Bi-GRU network. The final output of the s_i is represented by the following equation:

$$h_t = \vec{h}_t \oplus \overleftarrow{h}_t \tag{15}$$

then a multiple layer perceptron (MLP) is used to compute the scores of h_t as:

$$score_t = \sigma(W_y h_t + b_y) \tag{16}$$

TABLE 1. Details of the dataset.

Item	Number
Documents	2105
Clauses	11799
Emotions Causes	2167

Finally, $score_t$ is used to make a prediction for current clause c_t with softmax function.

E. REGULARIZATION

In order to eliminate the risk of overfitting caused by large number of parameters used in deep neural nets, we apply the regularization method proposed by [22] to the output of a dense layer, which revisits L2 regularization in the form of temporal activation regularization (TAR). It can be defined as:

$$\alpha L_2(h_t - h(t+1)) \tag{17}$$

where $L_2(\cdot) = ||\cdot||$, h_t is the output of the GRU at timestep t, and α is a scaling coefficient [22]. The TAR can penalize large changes in hidden state between timesteps, because of the prior it adds to minimize differences between states.

III. EXPERIMENTAL STUDY

A. DATASETS AND EVALUATION METRICS

Our experiments[1] are based on a simplified Chinese emotion cause corpus [8]. This corpus contains 2,105 documents from SINA city news and the details are shown in Table 1. Each document which has only one emotion word and one or more emotion causes is segmented into several clauses manually. The main task is to identify which clause contains the emotion cause [5].

We adopted commonly used evaluation metrics for this study, i.e., precision (P), recall (R), and F1 score. The evaluation metrics $P/R/F$ are defined as follows:

$$P = \frac{TP}{TP + FP} \tag{18}$$

$$R = \frac{TP}{TP + FN} \tag{19}$$

$$F1 = \frac{2 * P * R}{P + R} \tag{20}$$

where TP, FP, and FN represent true positives, false positives, and false negatives, respectively.

In the experiments, we randomly select 90% of the data set as training data and 10% as testing data. In order to obtain statistically credible results, we evaluate our method and baseline methods 25 times with different train/test splits.

B. BASELINES

In this research, we introduce six competitor models to study the contribution of each model component. These models are briefly described below:

- RB-CB-ML (Machine learning method trained from rule-based features and facts from a common-sense knowledge base): This methods proposed a multi-label approach to detect emotion causes [23]. It created two sets of linguistic patterns during feature extraction in order to capture the long-distance information to facilitate emotion-cause detection. This model has been tested in [5] who used a SVM with features extracted from the rules defined in [24].
- Word2Vec-SVM: It utilized a SVM classifier incorporated with word representations learned by a novel Word2vec model. The method presented a simplified variant of Noise Contrastive Estimation (NCE)for training the Word2Vec and used phrase-level making the Skip-gram model more expressive [15].It identified a large number of phrases with data-driven approach, and then considered the phrases as individual tokens during the training.
- ConvMS-Memnet: This approach considered emotion-cause identification as a reading comprehension task in QA [5]. It proposed a new deep memory slots to model the relation between a story and a query for QA system and to capture sequential information with the use of convolutional operations.
- CNN-MLP: In this model, we first model the emotion description and the clauses from the same document with convolutional network. Then we are inspired by the approach proposed by [25] to concatenate the sentence-level representation of emotion description with sentence vectors of clauses respectively, and these concatenated vectors are the input of MLP for emotion-cause detection.
- CNN-BiGRU-MLP_I: This model adds bi-directional gated recurrent units to get the information among clauses. Similar to the CNN+MLP method mentioned above, this model also uses CNN to get sentence representation and then concatenates vectors. Next, Bi-GRU is adopted to extract semantic information on sentence level. Finally MLP layer is implemented. This method is to prove the efficiency of Bi-GRU on the performance of emotion-cause extraction task.
- CNN-BiGRU-MLP_II: Compared with CNN-BiGRU-MLP_I, this model adopts a novel weighted pooling layer in convolutional neural network except using average pooling or max pooling. The motivation of this method is to evaluate the effectiveness of the weighted pooling on CNN model.

C. RESULT AND ANALYSIS

Table 2 shows the overall evaluation results. For Machine learning based method, RB-CB-ML and Word2Vec-SVM achieve relatively low F1 score compared with other deep learning based model. Although Word2Vec-SVM obtained word representation from the SINA news raw corpus, it also failed to perform well.

As to the deep learning approaches, ConvMS-Memnet

[1]The source code is available at `https://github.com/deardelia/ECextraction`

TABLE 2. Results compared with existing approaches.

Method	P(%)	R(%)	F1(%)
RB-CB-ML	59.21	53.07	55.97
Word2Vec-SVM	43.01	42.33	41.36
ConvMS-Memnet	70.76	68.38	69.55
CNN-MLP	58.96	57.82	58.34
CNN-BiGRU-MLP_I	66.92	64.08	65.47
CNN-BiGRU-MLP_II	69.73	66.84	68.25
HCS(Ours)	**73.88**	**71.54**	**72.69**

achieves F-measure of 69.55 and it far more outperforms all the machine learning methods mentioned above. The reason for its good performance can be explained by the structure of memory network and deep architecture, which can model both syntactic and semantic information and also consider an emotion lexicon.

Inspired by its success, other deep learning based models are implemented and the final version model HCS proposed in this paper reaches the best performance. Although CNN-MLP method achieves better than previous machine learning based model, it does not perform well. CNN-MLP only models features in a single clause rather than considering the relation with other clauses. Therefore, the utilization of Bi-GRU in CNN-BiGRU-MLP_I method proves the efficiency of semantic and syntactic information between clauses in the emotion-cause extraction task. CNN-BiGRU-MLP_I achieves F1 score of 65.47, which performs better than the simple CNN-MLP model by 7.13%. Then the application of weighted pooling method in CNN-BiGRU-MLP_II increases the F1 score by 2.78% compared with average pooling based CNN-BiGRU-MLP_II, as the weighted pooling mechanism can better focuses on the features related to emotion cause. HCS implementing an attention layer before CNN achieves the highest F1 score of 72.69%, which proves that the attention layer is effective to make the whole model focus on the emotion-related words.

D. EFFECTS OF ATTENTION MECHANISM

1) CONTENT ATTENTION

To gain better insights into our proposed model, we conduct other experiments to understand the influence brought out by content attention and local attention.

Attention layer attempts to measure the importance of each word in a clause according to the emotion description. In Figure 4, we choose to visualize the vector representation obtained from words in each clause included in a Fig. 1 after attention layer.

In this example, the cause of the emotion description ''too sad to faint'' is ''bad news about the death of her husband''. We adopt the widely used dimension reduction method t-Distributed Stochastic Neighbor Embedding (t-SNE) [26] to project word vectors into a two-dimensional space. In Fig. 4, each word is projected into a single point whose coordinate represents the weights projected on each direction. It can be seen that the words included in emotion cause, such as

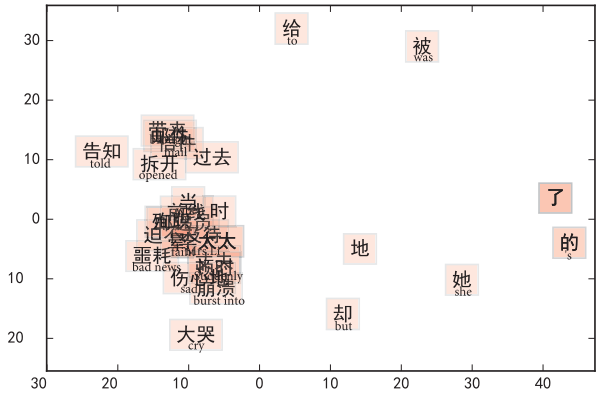

FIGURE 4. Attention visualization.

TABLE 3. Results compared with different position attention system.

Method	P(%)	R(%)	F1(%)
Strategy 1	74.43	**72.63**	**73.52**
Strategy 2	74.15	71.61	72.86
Strategy 3	**76.32**	69.99	73.02

''bad news'', ''death'', ''husband'' are close to the emotion description. This figure denotes that attention layer enables the cause-related words to gain more attention from emotion description.

2) POSITION ATTENTION

Table 3 shows the experiment results achieved by different position attention system. The Strategy1, 2, 3 represent the position attention strategies mentioned in part 3 respectively. From the table, we can notice that all of the position attention systems can contribute to the improvement of F1 score. Among them, Strategy 1 performs best, which proves the efficiency of position attention. The reason for the little improvement brought out by Strategy 2 might be that it only considers the distance between words and emotion description in the document level while ignores the importance of clause level.

3) VISUALIZATION OF ATTENTION MECHANISM

In order to get better understanding of the effects of attention mechanism, we visualize the attention weight of each words in a clauses and the results can be seen in Table 4. Among the table, it should be noticed that we consider the weighted pooling in CNN model as another kind of attention system, as it also figures out the weights of words.

E. EFFECTS OF ACTIVATION REGULARIZATION

Here we discuss the efficiency of activation regularization (TAR) adopted in our method with the ordinary L_2 approach. Figure 5 shows the changes of F1 score with the training batch. ''TAR_train'' and ''TAR_test'' represent the F1 score of training set and test set with the implementation of TAR, then ''R_train'' and''R_test'' are the F1 scores achieved by the use of ordinary regularization method.

TABLE 4. Visualization of word-level attention mechanism.

Words	Content attention	Position attention	Weighted pooling
却 (but)	0.09	0.06	0.05
被 (was)	0.07	0.06	0.03
告知 (told)	0.12	0.12	0.15
丈夫 (husband)	0.24	0.22	0.28
殉职(death)	0.25	0.24	0.28
的('s)	0.06	0.08	0.02
噩耗(bad news)	0.17	0.22	0.17

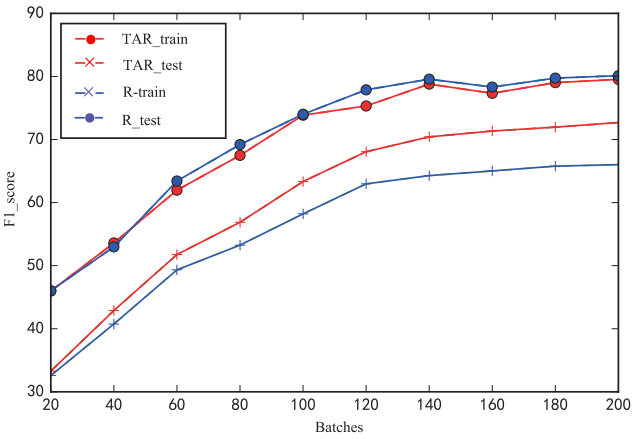

FIGURE 5. Comparison between different regularization methods.

According to the Fig. 5, although there is little difference between the F1 scores achieved by two regularization methods, the F1 score of test set using ordinary regularization is much lower than that adopting TAR, which suggests that TAR can reduce the risk of overfitting.

IV. RELATED WORK

Emotion cause extraction task is one of the fundamental challenges in the area of emotion analysis. Its goal is to reveal important information about what causes a certain emotion and why there is an emotion change [5].

In the currently existed studies in emotion recognition, deep convolutional neural networks (DCNNs) was widely adopted to learn the relevant and complex feature representation from short segments of data [27]. The original CNN model has been proven to exhibit better results than those of DNNs [28], who trained CNNs to classify and recognize EEG signals [29] according to the emotional states ''positive'' or ''negative''. CNNs require fewer parameters and are more suitable for inputting raw data because CNN layer does not need to disband the structure of data. Later on, as Recurrent Neural Networks (RNNs) have shown considerable success in many natural language processing (NLP) tasks, Lim et al. proposed a Speech Emotion Recognition (SER) method based on concatenated CNNs and RNNs without using any traditional hand-crafted features [30]. This method performed better compared to those conventional classification methods. In order to reduce the risk of overfitting, Li *et al.* [31] organized differential entropy features from different channels as two-dimensional maps to train a used hierarchical convolutional neural network (HCNN) to classify the positive, neutral, and negative emotion states. All the aforementioned work only focused on emotion recognition task rather than emotion-cause extraction.

As for emotion-cause extraction, Lee *et al.* first constructed a Chinese emotion cause corpus for designing automatic systems of emotion-cause detection and emotion classification [32]. Chen *et al.* [23] regarded this task as a case of cause detection and proposed a multi-label approach to detect emotion causes. They developed two sets of linguistic features for emotion-cause detection based on linguistic cues. Such rule-based method was popular in early studies. For example, Neviarouskaya and Aono [33] proposed a model which was based on the analysis of syntactic and dependency information from the parser in order to extract emotion causes automatically. Similarly, Gao *et al.* [7] also utilized a rule-based system which extracts the corresponding cause events in fine-grained emotions from the results of events, actions of agents and aspects of objects. Cheng *et al.* [34] focused on emotion-cause detection using a multiple-user structure (i.e., texts in a microblog are successively written by multiple users). Their experiments show that the current-subtweet-based emotion-cause detection is much more difficult than the original emotion-cause detection and needs further exploration.

Recently Gui *et al.* [8] constructed a new corpus using SINA city news and based on this corpus, they first used a convolution kernel based multi-kernel SVM method for event-driven emotion-cause extraction. Later on, they further verified the importance of analyzing clause and selecting suitable clause as the first and preliminary step towards emotion-cause detection. Based on this argument, they consider the emotion description as a question and clause as potential answers and then proposed a proposed a ConvMS-Memnet model to deal with the clause selection challenge [5] to this task.

V. CONCLUSION AND FUTURE WORK

Clauses are important sources for extracting emotion cause from documents. In this research we proposed a multiple level hierarchical network based clause selection (HCS) strategy to tackle this problem. The proposed hierarchical framework is able to detect emotion cause related clause from three levels: word level, phrase level and clause level, respectively. It can not only take semantic information between emotion

description and clause into consideration but also consider the relationships among clauses. Experimental results conducted on the publicly available Chinese emotion cause corpus demonstrate the effectiveness of the proposed model.

In our future work, it is interesting to explore more efficient ways to find the cause of emotion directly rather than output the clause which probably includes the emotion cause. Furthermore, we would like to apply our multiple level hierarchical network based clause selection model to other emotion analysis tasks.

REFERENCES

[1] L. Gui, R. Xu, Q. Lu, D. Wu, and Y. Zhou, "Emotion cause extraction, a challenging task with corpus construction," in *Proc. 5th Chin. Nat. Conf. Social Media Process.*, 2016, pp. 98–109.
[2] J. Gao, Y. Fu, Y.-G. Jiang, and X. Xue, "Frame-transformer emotion classification network," in *Proc. ACM Int. Conf. Multimedia Retr.*, 2017, pp. 78–83.
[3] W. Wei, Q. Jia, Y. Feng, and G. Chen, "Emotion recognition based on weighted fusion strategy of multichannel physiological signals," *Comput. Intell. Neurosci.*, vol. 2018, Jul. 2018, Art. no. 5296523.
[4] Y. Yao *et al.*, "Reader emotion prediction using concept and concept sequence features in news headlines," in *Proc. 15th Int. Conf. Intell. Text Process. Comput. Linguistics*, 2014, pp. 73–84.
[5] L. Gui, J. Hu, Y. He, R. Xu, Q. Lu, and J. Du, "A question answering approach for emotion cause extraction," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 1593–1602.
[6] W. Li and H. Xu, "Text-based emotion classification using emotion cause extraction," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1742–1749, 2014.
[7] K. Gao, H. Xu, and J. Wang, "A rule-based approach to emotion cause detection for chinese micro-blogs," *Expert Syst. Appl.*, vol. 42, no. 9, pp. 4517–4528, 2015.
[8] L. Gui, D. Wu, R. Xu, Q. Lu, and Y. Zhou, "Event-driven emotion cause extraction with corpus construction," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 1639–1649.
[9] C. Yang, T. Liu, L. Liu, X. Chen, and Z. Hao, "A personalized friend recommendation method combining network structure features and interaction information," in *Proc. 9th Int. Conf. Adv. Swarm Intell.*, 2018, pp. 267–274.
[10] S. Tafreshi and M. Diab, "Sentence and clause level emotion annotation, detection, and classification in a multi-genre corpus," in *Proc. 11th Int. Conf. Lang. Resour. Eval.*, 2018, pp. 1246–1251.
[11] F. Wu *et al.*, "Temporal interaction and causal influence in community-based question answering," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 10, pp. 2304–2317, Oct. 2017.
[12] J. Wang *et al.*, "Aspect sentiment classification with both word-level and clause-level attention networks," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, 2018, pp. 4439–4445.
[13] Y. Zhang, V. Zhong, D. Chen, G. Angeli, and C. D. Manning, "Position-aware attention and supervised data improve slot filling," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 35–45.
[14] F. Blain, V. Logacheva, and L. Specia, "Phrase-level segmentation and labelling of machine translation errors," in *Proc. 10th Int. Conf. Language Resour. Eval.*, 2016, pp. 2240–2245.
[15] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. 27th Annu. Conf. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.
[16] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 214–224.
[17] T. Shen, T. Zhou, G. Long, J. Jiang, S. Pan, and C. Zhang, "DiSAN: Directional self-attention network for RNN/CNN-free language understanding," in *Proc. 32nd AAAI Conf. Artif. Intell., Innov. Appl. Artif. Intell., 8th AAAI Symp. Educ. Adv. Artif. Intell.*, 2018, pp. 5446–5455.
[18] W. Yin, H. Schütze, B. Xiang, and B. Zhou, "ABCNN: Attention-based convolutional neural network for modeling sentence pairs," *Trans. Assoc. Comput. Linguistics* vol. 4, pp. 259–272, Jun. 2016.
[19] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder–decoder approaches," in *Proc. 8th Workshop Syntax, Semantics Struct. Stat. Transl.*, 2014, pp. 103–111.
[20] J. Li, X. Wang, Y. Zhao, and Y. Li, "Gated recurrent unit based acoustic modeling with future context," in *Proc. 19th Annu. Conf. Int. Speech Commun. Assoc.*, 2018, pp. 1788–1792.
[21] X. Luo, W. Zhou, W. Wang, Y. Zhu, and J. Deng, "Attention-based relation extraction with bidirectional gated recurrent unit and highway network in the analysis of geological data," *IEEE Access*, vol. 6, pp. 5705–5715, 2018.
[22] S. Merity, B. McCann, and R. Socher. (2017). "Revisiting activation regularization for language RNNs." [Online]. Available: https://arxiv.org/abs/1708.01009
[23] Y. Chen, S. Y. M. Lee, S. Li, and C.-R. Huang, "Emotion cause detection with linguistic constructions," in *Proc. 23rd Int. Conf. Comput. Linguistics*, 2010, pp. 179–187.
[24] S. Y. M. Lee, Y. Chen, and C.-R. Huang, "A text-driven rule-based system for emotion cause detection," in *Proc. NAACL HLT Workshop Comput. Approaches Anal. Gener. Emotion Text*, 2010, pp. 45–53.
[25] X. Zhou, B. Hu, Q. Chen, and X. Wang, "Recurrent convolutional neural network for answer selection in community question answering," *Neurocomputing*, vol. 274, pp. 8–18, Jan. 2018.
[26] N. Pezzotti, B. P. F. Lelieveldt, L. van der Maaten, T. Höllt, E. Eisemann, and A. Vilanova, "Approximated and user steerable tSNE for progressive visual analytics," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 7, pp. 1739–1752, Jul. 2017.
[27] W. Q. Zheng, J. S. Yu, and Y. X. Zou, "An experimental study of speech emotion recognition based on deep convolutional neural networks," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, Sep. 2015, pp. 827–831.
[28] M. Yanagimoto and C. Sugimoto, "Recognition of persisting emotional valence from EEG using convolutional neural networks," in *Proc. IEEE 9th IEEE Int. Workshop Comput. Intell. Appl.*, Nov. 2016, pp. 27–32.
[29] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "EEG-based emotion classification using deep belief networks," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2014, pp. 1–6.
[30] W. Lim, D. Jang, and T. Lee, "Speech emotion recognition using convolutional and recurrent neural networks," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA)*, Dec. 2016, pp. 1–4.
[31] J. Li, Z. Zhang, and H. He, "Hierarchical convolutional neural networks for EEG-based emotion recognition," *Cogn. Comput.*, vol. 10, no. 2, pp. 368–380, 2018.
[32] S. Y. M. Lee, Y. Chen, S. Li, and C. Huang, "Emotion cause events: Corpus construction and analysis," in *Proc. Int. Conf. Lang. Resour. Eval.*, 2010, pp. 1–8.
[33] A. Neviarouskaya and M. Aono, "Extracting causes of emotions from text," in *Proc. 6th Int. Joint Conf. Natural Lang. Process.*, 2013, pp. 932–936.
[34] X. Cheng, Y. Chen, B. Cheng, S. Li, and G. Zhou, "An emotion cause corpus for chinese microblogs with multiple-user structures," *ACM Trans. Asian Lang. Inf. Process.*, vol. 17, no. 1, pp. 6:1–6:19, 2017.

XINYI YU is currently pursuing the B.Sc. degree in computer science with the Honors College, Beihang University, China. Her research interests include machine learning, information retrieval, and natural language processing.

WENGE RONG received the B.Sc. degree from the Nanjing University of Science and Technology, China, in 1996, the M.Sc. degree from Queen Mary College, U.K., in 2003, and the Ph.D. degree from the University of Reading, U.K., in 2010. He is currently an Associate Professor with Beihang University, China. He has many years of working experience as a Senior Software Engineer in numerous research projects and commercial software products. His current research interests include machine learning, natural language processing, and information management.

ZHUO ZHANG is currently pursuing the B.Sc. degree with the Honors College, Beihang University, China. His main research interest includes natural language processing.

YUANXIN OUYANG received the B.Sc. and Ph.D. degrees from Beihang University, China, in 1997 and 2005, respectively, where she is currently an Associate Professor. Her research interests include recommender systems, data mining, social networks, and service computing.

ZHANG XIONG is currently a Professor with the School of Computer Science of Engineering, Beihang University, and the Director of the Advanced Computer Application Research Engineering Center, National Educational Ministry of China. He has published over 200 referred papers in international journals and conference proceedings. His research interests and publications span from smart cities, knowledge management, and information systems. He received the National Science and Technology Progress Award.