A Co-Attention Neural Network Model for Emotion Cause Analysis with Emotional Context Awareness

Xiangju Li[1] , **Kaisong Song**[2] , **Shi Feng**[1] , **Daling Wang**[1] , **Yifei Zhang**[1]
[1]School of Computer Science and Engineering, Northeastern University, Shenyang, China
lixiangju100@163.com, (fengshi, wangdaling, zhangyifei)@cse.neu.edu.cn
[2]Alibaba Group, Hangzhou, China kaisong.sks@alibaba-inc.com

Abstract

Emotion cause analysis has been a key topic in natural language processing. Existing methods ignore the contexts around the emotion word which can provide an emotion cause clue. Meanwhile, the clauses in a document play different roles on stimulating a certain emotion, depending on their content relevance. Therefore, we propose a co-attention neural network model for emotion cause analysis with emotional context awareness. The method encodes the clauses with a co-attention based bi-directional long short-term memory into high-level input representations, which are further fed into a convolutional layer for emotion cause analysis. Experimental results show that our approach outperforms the state-of-the-art baseline methods.

1 Introduction

Emotion cause analysis (ECA) aims to identify the reasons behind a certain emotion expressed in text. Compared with emotion classification (Gao et al., 2013; Song et al., 2016, 2017; Feng et al., 2018), ECA is more challenging because it requires deep comprehension of text semantics and then extracts the cause of emotions accurately. Recently, Gui et al. (2017) considered ECA as a question answer problem and proposed a method which took a query (i.e., a single emotion word) as the input of a QA system. However, their method ignores an important fact that the context around the emotion word details the emotions and serves as an emotion cause clue. Therefore, the emotion clause consisting of the emotion word and its context should be treated as a whole unit for query.

Example 1 *In 2016* (c_1)*, Lakers' Kobe announced his retirement* (c_2)*. In the same year* (c_3)*, Spurs' Duncan also retired* (c_4)*. The Spurs' athletes and Duncan's fans are frustrated* (c_5)*.*

For example, a microblog user published a post in Twitter (see Example 1). From Example 1, we can find that the clause c_5 is the emotion clause which contains the emotion word "*frustrated*" and the clause c_4 is the cause clause of the emotion. However, it is difficult for previous works (Gui et al., 2017) to extract the expected emotion causes by only taking the emotion word "*frustrated*" as a query. In other words, the retirement of Kobe can also stimulate "*frustrated*" emotion, but it is not the cause of the clause c_5. Therefore, the context around the emotion word in the same emotion clause can serve as a useful emotion cause clue for the ECA task. In addition, the words shared by the emotion cause clause and the emotion clause (e.g., "*Spurs*" and "*Duncan*") usually reflect stronger emotion-cause associations and should be paid more attention when modeling.

In this paper, we propose a novel, extensive and effective Co-Attention Neural Network method (CANN) for this ECA task. The method incorporates the contextual information by a bi-directional long short-term memory encoder. Then, an attention matrix is constructed to model the mutual impacts of each candidate cause clause and the whole emotion clause. Meanwhile, the above clauses are paid different attention based on the transformed attention matrix. Finally, a convolutional neural network is utilized for classification. To sum up, the main contributions of our paper are three-fold: (1) We present a novel Co-Attention Neural Network model (CANN), which identifies the emotion cause from subjective texts by combining a co-attention mechanism, bi-directional long short-term memory units and convolutional neural network into a unified framework. (2) A co-attention mechanism is designed to capture the mutual impacts between each candidate cause clause and the emotion clause for enhancing the input representations of a neural network model. (3) Our

Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 4752–4757
Brussels, Belgium, October 31 - November 4, 2018. ©2018 Association for Computational Linguistics

comparative results on a public benchmark dataset show that CANN outperforms several state-of-the-art methods, which verifies the effectiveness of our proposed method.

2 Related Work

A variety of learning methods have been applied to emotion cause analysis and achieved plausible improvements since Lee et al. (2010) first gave the formal definition of this task and constructed a public available Chinese emotion cause annotated corpus based on manual annotation scheme. Chen et al. (2010) developed two sets of linguistic features based on linguistic cues and proposed a rule-based approach. Similarly, Gui et al. (2014) also proposed a rule based emotion cause detection method by designing 25 manually complied rules for this task and Gao et al. (2015) extended this method to deal with informal text such as microblog. Ghazi et al. (2015) built a Conditional Random Fields (CRF) learner to detect the emotion stimulus. However, this model is based on a simplistic assumption that emotion cause and emotion word co-exist in the same sentence, which limits its application scenario. More recently, Gui et al. (2016b) proposed a multi-kernel based method trained from a public SINA city news ECA Corpora. However, this method is still heavily dependent on the design of effective features.

With the apparent success of neural network methods and attention mechanisms, Gui et al. (2017) proposed a novel deep neural networks model which considers emotion cause identification as a question answering task. However, their method only takes a single emotion word as the query, which ignores important emotion-cause clues. Meanwhile, their method fails to utilize the mutual influences between each candidate cause clause and the emotion clause, which will depress the model performance. To address these issues, we propose a novel CANN model which not only considers the whole emotion clause as a query but also models the mutual influences between the candidate cause clause and the whole emotion clause via co-attention mechanism.

3 Methodology

In this section, we will first give the formal definition of the emotion cause analysis task. Then, we introduce our proposed method in detail.

3.1 The Definition of Emotion Cause Analysis

Given a document $D = \{c_1, c_2, \cdots, c_n\}$ about an emotion cause event, the goal of this ECA task is to identify which clause stimulates the emotion expression. For convenience, we suppose each document has a unique emotion clause (containing a single emotion word) and at least a corresponding emotion cause clause. Previous work considers emotion cause analysis as a question answering task by taking a single emotion word as a query to retrieve the emotion cause (Gui et al., 2017). This is not applicable to our observation that context around the emotion word in the same clause can serve as an important emotion-cause clue. Here, we are intended to take the whole emotion clause (i.e., the emotion word and its context words) as a query. Therefore, the emotion cause analysis task can be formulated as below:

$$P_x = \text{emotion-cause}\left(c^q, c^a\right)$$

where c^q is an emotion clause (i.e., query) and c^a is a candidate emotion cause clause (i.e., candidate answer)[1]. The probability P_x of identifying the clause c^a as the emotion cause of the emotion clause c^q is computed by the emotion-cause function ($x = 1$ if c^a stimulates the emotion expressed in the c^q, or $x = 0$ otherwise).

3.2 Co-Attention Neural Network for Emotion Cause Analysis

In this section, we propose a Co-Attention Neural Network (CANN) method, which contains four different layers: (1) an embedding layer for the input words; (2) a bi-directional long short-term memory (Bi-LSTM) layer; (3) a co-attention layer for generating the attention maps of the emotion clause and the candidate cause clause; and (4) a convolutional neural network layer to generate a probability distribution over the output classes. Figure 1 illustrates the details of our CANN.

(1) The embedding layer maps each word w_i into a d-dimensional vector $v_i \in \mathbb{R}^d$. Then, the clause c^q can be represented as a feature map $E^q = [v_1^q, v_2^q, ..., v_n^q]$. Symmetrically, we can also produce E^a for the emotion cause clause c^a.

(2) Bi-LSTM layer aims to encode contextual information with a Bi-LSTM (Wang et al., 2016). Let v_t^q be the t-th word vectors of the clause c^q. Each contextual embedding at time-step t of c^q can be formulated as $h_t^q = concat(\overrightarrow{h}_t^q, \overleftarrow{h}_t^q)$, where

[1]Note that the emotion clause c^q and the emotion cause clause c^a can be the same clause.

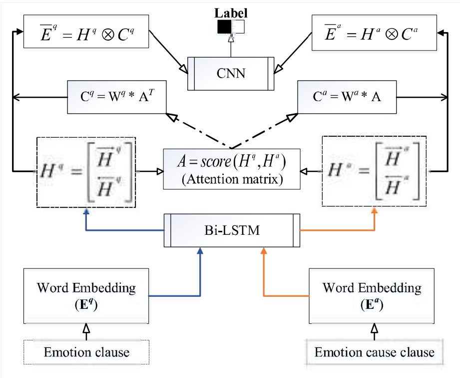

Figure 1: CANN architecture for ECA task.

$concat(\cdot,\cdot)$ is a concatenation operation, $\overrightarrow{h}_t^q = \overrightarrow{LSTM}(\overrightarrow{h}_{t-1}^q, v_t^q)$ and $\overleftarrow{h}_t^q = \overleftarrow{LSTM}(\overleftarrow{h}_{t+1}^q, v_t^q)$ are hidden states produced by a forward LSTM and a backward LSTM, respectively. Finally, a context embedding matrix $H^q \in \mathbb{R}^{n\times 2h}$ of the clause c^q can be obtained by concatenating all the context embeddings, where h is the number of hidden units. Similarly, we can also generate H^a for the emotion cause clause c^a.

(3) Co-attention layer contains three steps: construction of attention matrix, construction of attention feature maps and input transformation.

Firstly, we construct an attention matrix $A \in \mathbb{R}^{n\times n}$ which models mutual influences between the c^q and c^a via a measurement function: $A = score\big(H^q, H^a\big)$. Element $A_{i,j}$ represents the semantic similarity between the i-th word of the clause c^q and the j-th word of the c^a clause[2].

Next, two attention feature maps $C^q \in \mathbb{R}^{2h\times n}$ and $C^a \in \mathbb{R}^{2h\times n}$ can be derived from linearly transformed attention matrix A: $C^q = W^q \cdot A^\top$ and $C^a = W^a \cdot A$, where $W^q \in \mathbb{R}^{2h\times n}$ and $W^a \in \mathbb{R}^{2h\times n}$ are weight matrices.

Finally, we apply the attention feature maps to the original representations of c^q and c^a respectively: $\bar{E}^q = H^q \otimes C^q$, $\bar{E}^a = H^a \otimes C^a$.[3] They will be fed into CNN layer for future operations.

(4) In CNN layer, we applied three convolution filters with lengths 2, 3 and 4, each of which is followed by a max pooling operation to select most informative features. Two high-level feature vectors z^q and z^a are generated and then concatenated as the inputs of a fully interconnected softmax layer before classification.

[2] function $score$ is the cosine distance

[3] $\otimes$ denotes element-wise product.

Model Training: We use cross-entropy error between gold distribution and predicted distribution as the loss function. We use back-propagation to calculate the gradients of all the trainable parameters and update them with Adam (Kingma and Ba, 2015).

4 Experiments and Results

4.1 Dataset and Setup

Our experiments are conducted based on a Chinese emotion cause dataset[4]. The statistics are presented in Table 1. We evaluate the performance by precision (P), recall (R), and F-measure (F), which are commonly used evaluation metrics (Gui et al., 2016a). We set the number hidden units $h = 50$, the number of filters as $N = 100$, the dimension of word embedding $d = 30$, the initial learning rate $lr = 10^{-4}$, the dropout rate $p = 0.5$ and the batch size as 50 according to the best F value on the development set which is set aside 10% from the training set.

Item	Number	Item	Number
#Documents	2,105	#Cause$_1$	2,046
#Causes	2,167	#Cause$_2$	56
#Clauses	11,799	#Cause$_3$	3

Table 1: Statistics of the experimental dataset we used. Cause$_1$, Cause$_2$ and Cause$_3$ represent the documents with 1, 2 and 3 cause clauses, respectively.

4.2 Comparison of Different Methods

We compare CANN with some traditional and advanced baselines in Table 2, and separate results into three groups: (1) rule based or common-sense based methods; (2) machine learning methods; (3) deep neural network methods.

(1) **RB** is a rule based method, which generalizes two sets of linguistic rules for emotion cause analysis (Lee et al., 2010). **CB** is a knowledge based method (Russo et al., 2011), which uses the Chinese Emotion Cognition Lexicon (Xu et al., 2013) as the common-sense knowledge base.

(2) **RB+CB+ML** is a support vector machine (SVM) classifier trained on features including rules (Lee et al., 2010) and the Chinese Emotion Cognition Lexicon (Xu et al., 2013). **SVM** is a SVM classifier trained on unigrams, bigrams and

[4] Dataset: http://hlt.hitsz.edu.cn/?page%20id=694

Methods	*P*	*R*	*F*
RB$^{\diamond}$	**0.6747**	0.4287	0.5243
CB$^{\diamond}$	0.2672	**0.7130**	0.3887
RB + CB$^{\diamond}$	0.5435	0.5307	**0.5370**
RB + CB + ML$^{\diamond}$	0.5921	0.5307	0.5597
SVM$^{\diamond}$	0.4200	0.4375	0.4285
Word2vec$^{\diamond}$	0.4301	0.4233	0.4136
Multi-kernel$^{\diamond}$	**0.6588**	**0.6927**	**0.6752**
CNN	0.5307	0.6427	0.5807
ConvMS-Memnet$^{\diamond}$	0.7076	0.6838	0.6955
CANN	**0.7721**	**0.6891**	**0.7266**

Table 2: Comparison among different methods. The results with superscript $\diamond$ are reported in (Gui et al., 2017). The best results in each group are highlighted.

trigrams features. **Word2vec** uses word representations obtained by Word2vec (Mikolov et al., 2013b) as features and then trains a SVM classifier. **Multi-kernel** (Gui et al., 2016a) uses the multi-kernel method to identify the emotion cause.

(3) **ConvMS-Memnet** is the current state-of-the-art method using the convolutional multiple-slot deep memory network to identify the emotion causes (Gui et al., 2017). **CNN** is the basic convolutional neural network proposed by (Kim, 2014). **CANN** is our co-attention neural network method.

From Group 1 in Table 2, we can observe that RB+CB outperforms RB and CB methods in F, which verifies that RB and CB are complementary to improve the model performance. In Group 2, Multi-kernel performs the best among the machine learning methods because it considers the structured context information, which plays an important role in this task. In Group 3, the best F value is achieved by our proposed CANN, which outperforms the state-of-the-art method ConvMS-Memnet by 3.11%. Meanwhile, the optimal values on P and R are also generated by CANN method. The results imply that considering context around the emotion word is helpful to capture useful emotion clues and detail context information. We can also find that CANN outperforms CNN method by 14.33% in F, which indicates that our co-attention mechanism has a significant positive effect for emotion cause identification by capturing the mutual impacts between the emotion clause (i.e., query) and each candidate cause clause (i.e., candidate answer).

4.3 Effects of Different CANN Components

We further validate the different roles of model components by comparing several sub-networks based on our proposed architectures: **Full** is our proposed CANN method which considers the whole emotion clause as a query unit. **Full-BCA** does not contain the Bi-LSTM encoder and the co-attention mechanism. **Full-CA** does not consider the co-attention mechanism. **Full-EA** only pays attention to the candidate cause clause.

Methods	*P*	*R*	*F*
Full-BCA	0.5307	0.6427	0.5807
Full-CA	0.7209	0.5405	0.6147
Full-EA	0.7714	0.6596	0.7100
Full	**0.7721**	**0.6891**	**0.7266**

Table 3: Effects of different CANN components.

From Table 3, we can find that all the partial configured models cannot compete with the **Full** model indicating that the co-attention mechanism is effective for the emotion cause analysis task. Meanwhile, the F value is improved remarkably when co-attention mechanism is considered. **Full-CA** performs better than **Full-BCA**, which verifies the effectiveness of Bi-LSTM in clause presentation. We also notice that **Full** gains nearly 12% improvements in F compared with **Full-CA** , which indicates that both emotion clause and emotion contribute to acquiring better performance.

4.4 Effects of Emotion Word and Context Words around Emotion Word

In this section, we further study the effects of context words around emotion word in the same emotion clause on the classification performance. **Full** is our proposed CANN method which considers the whole emotion clause as a query unit. **Full-emotion** does not consider the emotion word. **Full-context** only considers emotion word but ignores its context words. We compare above three different configurations of CANN in Table 4.

Methods	*P*	*R*	*F*
Full-emotion	0.6514	0.7203	0.6827
Full-context	0.5534	0.6815	0.6095
Full	**0.6939**	**0.7601**	**0.7240**

Table 4: Effects of emotion word and context words around emotion word.

As shown in Table 4, it is clear that the partial configurations cannot compete with the **Full** model indicating both emotion word and its context words are helpful. Specifically, emotion word can provide emotion clues directly and its context words can provide richer emotional details, both of which contribute to capturing intrinsic emotion causes. **Full-context** performs slightly better than **Full-emotion** because context words reflect the "skeleton" of emotion expression directly, such as "I feel so much _" where the whitespace "_" can be any emotion word (e.g., happy or sad). This verifies the advantages of context words.

4.5 Effects of Initialization Methods

We further study the influences of different initialization ways by applying different word embeddings into the input layer of our CANN method. **Random** initializes word embeddings randomly from a truncated standard normal distribution. **CBOW** and **Skip** represent the word embeddings which are pre-trained respectively by CBOW and Skip-Gram models (Mikolov et al., 2013a) based on the training set and then applied to parameter initialization.

Methods	*P*	*R*	*F*
Random	0.6519	0.5650	0.6022
CBOW	0.7207	**0.7245**	0.7226
Skip	**0.7721**	0.6891	**0.7266**

Table 5: Effects of different initialization methods.

As shown in Table 5, pre-trained initialization methods perform better than random initialization, which indicates that CANN is sensitive to the initial settings of the parameters. Meanwhile, **Skip** method performs slightly better than **CBOW** method by achieving best results in metrics P and F. Therefore, we adopt the pre-trained word embeddings generated by **Skip** method to initialize all the word vectors of our CANN in this work.

4.6 Case Study

In order to validate whether our CANN can focus on the words which describe the emotion causes. We select two clauses ($\mathbf{c_2}$ and $\mathbf{c_4}$) from Example 1 and visualize the attention feature map C^a in Figure 2. The horizontal coordinate represents the different dimensions of each word representation obtained by Bi-LSTM. The vertical coordinate represents the words of the clause. And the color depth represents different degrees of the attention weights of each word in emotion cause clause.

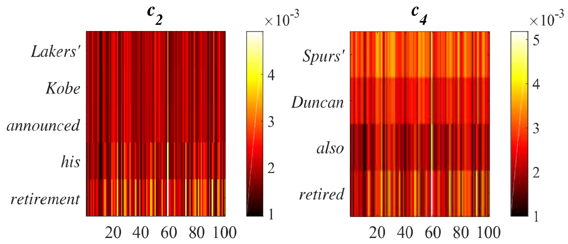

Figure 2: Visualization of Attention.

From Figure 2, we can easily find the word "*retirement*" in the clause $\mathbf{c_2}$ has much higher attention weights than other words in this emotion cause clause, which indicates that the cause is closely related to the word "*retirement*". In the clause $\mathbf{c_4}$, the words "*Spurs'*" and "*Duncan*" also have high attention weights expect "*retired*", which shows that the emotion cause is not only related to "*retirement*" and "*retired*" but also "*Spurs'*" and "*Duncan*". Above all, clause $\mathbf{c_4}$ is more likely to be identified as the emotion cause clause according to our method, which is consistent with our observation. We observe that words like *"Lakers'"*, *"his"* and *"also"* have low attention weights and they are irrelevant to the emotion word *"frustrated"*. These again verify the effectiveness of our proposed attention mechanism on the emotion cause analysis task.

5 Conclusions

This paper proposes an effective Co-Attention Neural Network to identify the emotion causes. The mutual impacts between the emotion clause and each candidate clause are modeled as an attention matrix which further transforms original inputs into high-level input features of a constitutional neural network for classification. Experimental results conducted on a real-word dataset shows that our method achieves substantial improvements against "state-of-the-art" baselines.

Acknowledgments

The work was supported by the National Key R&D Program of China under grant 2018YFB1004700, and National Natural Science Foundation of China (61772122, 61872074).

References

Ying Chen, Sophia Yat Mei Lee, Shoushan Li, and Churen Huang. 2010. Emotion cause detection with linguistic constructions. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 179–187.

Shi Feng, Yaqi Wang, Kaisong Song, Daling Wang, and Ge Yu. 2018. Detecting multiple coexisting emotions in microblogs with convolutional neural networks. *Cognitive Computation*, 10(1):136–155.

Kai Gao, Hua Xu, and Jiushuo Wang. 2015. Emotion cause detection for chinese micro-blogs based on e-cocc model. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 3–14.

Wei Gao, Shoushan Li, Sophia Yat Mei Lee, Guodong Zhou, and Churen Huang. 2013. Joint learning on sentiment and emotion classification. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1505–1508.

Diman Ghazi, Diana Inkpen, and Stan Szpakowicz. 2015. Detecting emotion stimuli in emotion-bearing sentences. In *Computational Linguistics and Intelligent Text Processing - 16th International Conference*, pages 152–165.

Lin Gui, Jiannan Hu, Yulan He, Ruifeng Xu, Qin Lu, and Jiachen Du. 2017. A question answering approach to emotion cause extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1953–1602.

Lin Gui, Dongyin Wu, Ruifeng Xu, Qin Lu, and Yu Zhou. 2016a. Event-driven emotion cause extraction with corpus construction. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1639–1649.

Lin Gui, Ruifeng Xu, Qin Lu, Dongyin Wu, and Yu Zhou. 2016b. Emotion cause extraction, a challenging task with corpus construction. In *Chinese National Conference on Social Media Processing*, pages 98–109.

Lin Gui, Li Yuan, Ruifeng Xu, Bin Liu, Qin Lu, and Yu Zhou. 2014. Emotion cause detection with linguistic construction in chinese weibo text. In *Natural Language Processing and Chinese Computing*, pages 457–464.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1746–1751.

Diederik P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, pages 1–15.

Sophia Yat Mei Lee, Ying Chen, and Churen Huang. 2010. A text-driven rule-based system for emotion cause detection. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 45–53.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Irene Russo, Tommaso Caselli, Francesco Rubino, Ester Boldrini, and Patricio Martínez-Barco. 2011. Emocause: an easy-adaptable approach to emotion cause contexts. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 153–160.

Kaisong Song, Ling Chen, Wei Gao, Shi Feng, Daling Wang, and Chengqi Zhang. 2016. Persentiment: A personalized sentiment classification system for microblog users. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 255–258.

Kaisong Song, Wei Gao, Shi Feng, Daling Wang, Kamfai Wong, and Chengqi Zhang. 2017. Recommendation vs sentiment analysis: A text-driven latent factor model for rating prediction with cold-start awareness. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pages 2744–2750.

Yang Wang, Shi Feng, Daling Wang, Yifei Zhang, and Ge Yu. 2016. Context-aware chinese microblog sentiment classification with bidirectional lstm. In *Asia-Pacific Web Conference*, pages 594–606.

Ruifeng Xu, Chengtian Zou, Yanzhen Zheng, Xu Jun, Lin Gui, Bin Liu, and Xiaolong Wang. 2013. A new emotion dictionary based on the distinguish of emotion expression and emotion cognition. *Journal of Chinese Information Processing*, 27(6):82–90.