Adversarial Training based Cross-lingual Emotion Cause Extraction

Hongyu Yan[1], Qinghong Gao[1], Jiachen Du[⋆1], Binyang Li[2] and Ruifeng Xu[⋆⋆1]

1. Harbin Institute of Technology,Shenzhen
2. School of Information Science and Technology, University of International Relations
hongyu_yan_235@163.com, gaoqinghong1994@gmail.com, jacobvan199165@gmail.com, byli@uir.edu.cn, xuruifeng@hit.edu.cn

Abstract. Emotion cause extraction (ECA) aims to identify the reasons behind a certain emotion expression in a text. It is a key topic in natural language processing. Existing methods relies on high-quality emotion resources and focuses on only one language. However, the public annotated corpora is fairly rare. Therefore, we propose an adversarial training based cross-lingual emotion cause extraction approach to leverage the semantic and emotion knowledge in a resource-abundant language (source language) for ECA in a resource-scarce language (target language). Instead of large-scale parallel corpora, we capture task-related but language-irrelevant features only on a small-scale Chinese corpora and an English corpora. In addition, an attention mechanism based on position and emotion expression information is designed to obtain the key parts of the clause devoting to ECA. Our proposed approach could capture rich semantic and emotion information in ECA learning process. It is demonstrated that our method can achieve better performance than the state-of-the-art results.

1 Introduction

With the flourishing development of Internet, emotion analysis has attracted much attention in field of Natural Language Processing (NLP). Most of previous researches focus on emotion classification or emotion detection. However, underlying information such as the cause of emotion needs to be extracted and analyzed in many real word applications which provides crucial information for applications ranging from economic forecast and public opinion mining to product design.

Example 1: *Because he was just attacked with a torrent of abuse online* (c_1). *He felt deeply angry* (c_2).

Emotion cause extraction (ECA) aims to identify the reason behind a certain emotion expression automatically. As shown in Example 1, "angry" is an emotion expression and the cause of "angry" is c_1. It is more challenging compared to emotion classification because it requires deeper understanding on text semantic.

⋆ Corresponding author

⋆⋆ Corresponding author

Existing approaches to identify emotion cause mainly concentrate on rule based methods [1, 2] and machine learning algorithms [3] which ignore the trigger relations between emotion expression and emotion cause. Recently, Gui [4] considered emotion cause extraction as a question answering task to capture the semantic relations between emotion expression and emotion cause. However, above researches fasten on only one language to identify the emotion cause extraction as well as public annotated corpora are rare and imbalanced in different languages, which results in the lack of utilizing abundant information from diverse languages.

In this paper, we present a cross-lingual based approach on emotion cause detection to leverage resources in a resource-rich language (such as English) to improve the performance in a resource-scare language (such as Chinese) by making most of the cross-lingual transfer knowledge. The traditional cross-lingual approaches are based on translated resources such as bilingual dictionary and parallel corpora or employ machine translation (MT) systems to translate corpora in the source language into the target language [5]. These methods are restricted because of the gap between the source language and the target language as well as the accuracy of MT systems. Thus, to overcome these issues, inspired by [6, 7], we propose an adversarial training based cross-lingual architecture (ATCL-ECA) to model cross-lingual semantic and emotion information between two languages without relying on extra parallel corpora.

The major contributions of our work can be summarized as follows:

- We propose a cross-lingual approach (ATCL-ECA) to learn language-irrelevant but task-related (LI-TR) information for emotion cause extraction.
- Instead of large-scale parallel corpora, adversarial training based method is conducted on the small-scale labeled Chinese corpora from [3] and English corpora from ECA [8]. In addition, an attention mechanism based on position and emotion expression information is designed to obtain the key parts of the clause devoting to ECA.
- It is demonstrated that our proposed model can capture cross-lingual semantic information to bridge the gap between two languages effectively for ECA task and outperforms the state-of-the-art approach on a public benchmark dataset.

2 Related Work

In this section, we review the literature related to this paper from two perspectives: emotion cause extraction and cross-lingual method.

2.1 Emotion Cause Extraction

With further researching on emotion analysis, the emotion cause corresponding to an emotion expression becomes noteworthy. Emotion cause extraction could reveal the cause information which triggers the emotion expression. Lee [9] first gave the formal definition of this task. They manually constructed a Chinese emotion cause corpora from the Academia Sinica Balanced Chinese Corpus. Based on above researches, Chen [2] put this task down to multi-label classification which can capture the long distance information basing on rule-based and semantic features. Other than these rule based methods, Ghazi [10] employed Conditional Random Fields (CRFs) to identify emotion causes. However, this study requires emotion cause and emotion expression in the same clause. Recently, Gui [3] proposed a multi-kernel based method to detect the emotion cause on a public Chinese emotion cause corpora. Whereas, it fairly depended the design of features. Inspired by the neural network, Gui [4] considered this task as a question learning problem by constructing a convolutional memory network. However, since public annotated corpora on this task is heavily rare, existing researches almost focus on only one language which ignore the cross-lingual knowledge obtaining from abundant resources to other languages.

2.2 Cross-lingual Emotion Analysis

The goal of cross-lingual emotion analysis is to bridge the gap between the source language and target language. Machine translation (MT) or parallel corpora based approaches are usually employed to solve this problem. Machine translation based methods use MT systems to project training data into the target language or test data into source language. Wan [11] proposed a co-training method which depends on machine translation. They first translated Chinese testing data into English, and English training data into Chinese. After that, they performed training and testing data into two independent views, i.e., English view and Chinese view. Li [12] opted the samples in the source language that were similar to those in the target language to minish the gap between two languages. With the development of deep learning, most of scholars adopt it to learn cross-lingual representations with parallel corpora. Zhou [13] proposed a cross-lingual representation learning model which simultaneously learn both the word and document representations on these two languages. Besides, Zhou [14] use a hierarchical attention model to jointly train with a bidirectional LSTM to learn representation. However, such approaches requires large-scale task-related parallel corpora. To solve above issues, we proposed a cross-lingual architecture via adversarial training to detect the emotion cause. It can not only capture the cross-lingual semantic information between two languages but also be conducted on only small-scale corpora.

3 Model

3.1 Task Definition

The formal definition of emotion cause extraction is proposed in [3]. Given a document $Doc = \{c_1, c_2, \cdots, c_n\}$ consisting of an emotion expression e and n clauses, which is a passage about an emotion event. Moreover, each clause $c =$

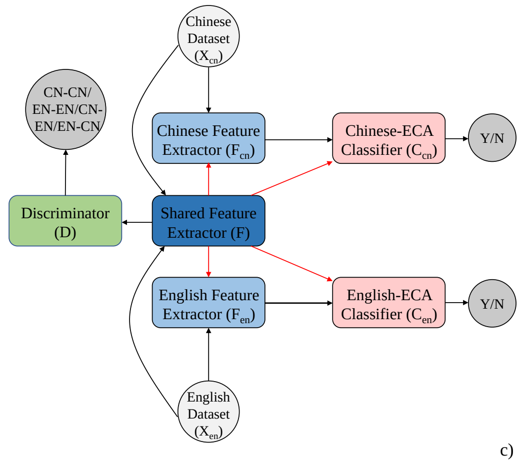

Fig. 1. Adversarial training based cross-lingual ECA architecture (ATCL-ECA).

$\{w_1, w_2, \cdots, w_k\}$ consists of k words. The goal of this task is to identify the emotion cause clause corresponding to the emotion expression. When dealing with each document, we map each word into a low dimensional and continuous vector, which is known as word embedding [15]. All the word vectors are stacked in a word embedding matrix $L = R^{dim} \times \|V\|$, where dim is the dimension of word vector and V is the vocabulary size.

3.2 Adversarial Training based Cross-lingual ECA Model

The overall architecture of Adversarial Training based Cross-lingual ECA Model is illustrated in Figure 1, which contains three components, i.e., feature extractor, ECA classifier and discriminator. We first use feature extractor to obtain the contextual information from different corpora, which consists of Chinese feature extractor F_{cn}, English feature extractor F_{en} and shared feature extractor F to acquire information from Chinese corpora and English corpora as well as shared semantic information between these two corpora. Then, to identify emotion cause corresponding to the emotion expression, we construct two classifiers C_{cn} and C_{en} on Chinese and English corpora respectively. It is well know that there is common knowledge notwithstanding the languages are diverse. To make most of this information, we concatenate the outputs of F_{cn} and the Chinese outputs from F then feed them into C_{cn}. Analogously, we feed the connection of outputs from F_{en} and the Chinese outputs from F then feed them into C_{en}. Besides, discriminator D is used to identify the language category of samples from F. The details of these portions are described in the following subsections.

Feature Extractor Normally, emotion expressions and emotion causes are expressed via phrases or sentences rather than only one word. Meanwhile, the same word in different contexts could convey different meanings. To incorporate rich contextual semantic information, we leverage Recurrent Neural Network with Gated Recurrent Unit (GRU) [16] to extract the word sequence features and long term. GRU instead of LSTM are exploited as the former has fewer number of parameters to tune, which contains only update gate z and reset gate r to control the flow of information. For each time step t, GRU first calculates the update gates z_t and reset gate r_t. Separately, the gate z_t is designed to regulate the degree of units updated to ensure that there are dependencies at every moment and the gate r_t determines the amount of selection for the previous state. The final hidden state h_t and candidate hidden state $\tilde{h_t}$ are obtained as follows:

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

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \tag{2}$$

$$\tilde{h}_t = tanh(W_h x_t + r_t \odot (U_h h_{t-1}) + b_h) \tag{3}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{4}$$

where x_t is word embedding of word w at time step t. $\sigma(\cdot)$ and $tanh(\cdot)$ are sigmoid function and hyperbolic tangent function separately. W_z, W_r, U_z, U_r, W_h and U_h are weight matrixes. $\odot$ is an Element-Wise Multiplication.

To capture the semantic features on Chinese corpora, English corpora and the shared knowledge from two languages, we use three feature extractors: F_{cn}, F_{en} and F. For each feature extractor, we adopt bidirectional GRU (Bi-GRU) to incorporate the past and future contextual information. The Bi-GRU comprises the forward GRU $\overrightarrow{f}$ which reads the sentence from left to right to learn the historical information and the backward GRU $\overleftarrow{f}$ which reads from right to left to obtain the future information.

$$\overrightarrow{h_{it}} = \overrightarrow{GRU_w}(x_{it}), t \in [1, k] \tag{5}$$

$$\overleftarrow{h_{it}} = \overleftarrow{GRU_w}(x_{it}), t \in [k, 1] \tag{6}$$

Then we concatenate the forward hidden state $\overrightarrow{h_{it}}$ and backward hidden state $\overleftarrow{h_{it}}$, i.e., $h_{it} = [\overrightarrow{h_{it}}, \overleftarrow{h_{it}}]$, which summarizes the information of the whole sentence around the word w.

ECA Classifier For emotion cause extraction task, not only semantic information of context is important, but also the relations between emotion expressions and emotion causes are salient features. Inspired by [17], to capture the semantic relations between emotion expressions and contexts, we utilize GRU-CNN attention based model (GRU-CNN-A). As shown in Figure 2, we first use Bi-GRU to obtain the contextual information in order that each word contains global semantic information, i.e., $c' = f(c)$, where f represents Bi-GRU. It is well acknowledged that not all words contribute equally to contexts. Hence, in order to capture the crucial components of the context, we employ attention mechanism.

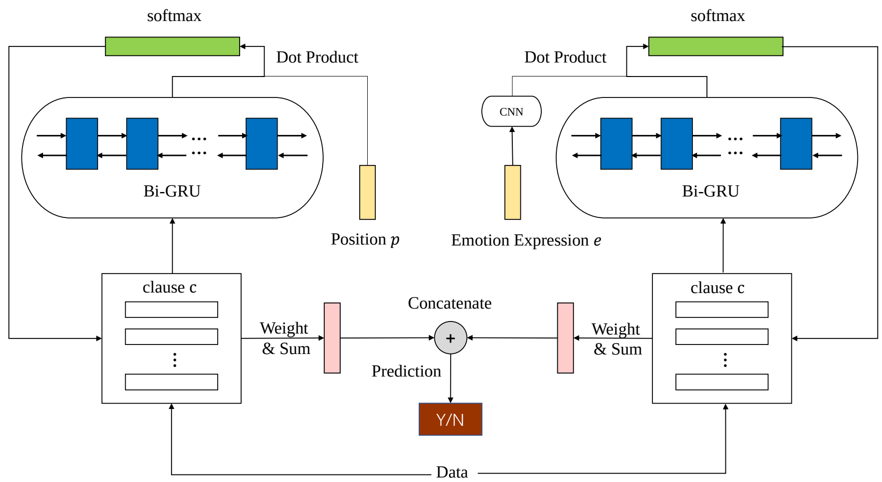

Fig. 2. GRU-CNN attention based emotion cause identification model.

Moreover, it can prompt the model to pay more or less attention to individual words or sentences.

Obviously, the emotion expressions are triggered by specific emotion cause events. Normally, these events are closed to the emotion expressions. Table 1 illustrates the distribution of cause positions and we can see most of emotion cause clauses abut the emotion expressions. Therefore, apart from emotion expressions themselves the distance between current clause and emotion words play core roles in emotion cause identification. In this paper, we combine position information p and emotion expression e to acquire attention of a word. Firstly, we obtain the word attention by the relative distance p between current clause and emotion expression e. Specifically,

$$m_p = c' \cdot W^p \cdot p \tag{7}$$

$$\alpha_j^p = \frac{exp(m_p^j)}{\sum_{k=1}^{M} exp(m_k^e)} \tag{8}$$

where W_p is the weight matrixes. α_j^p represents the importance of word in position j corresponding to the position vector p. That is, we make matrix multiplication between the output of Bi-GRU c' and the position vector p as well as get a normalized importance weight α_j^p through a softmax function. Similarly, the importance of words in contexts is also corresponding to the emotion expression e.

$$m_e = c' \cdot W^e \cdot CNN(e) \tag{9}$$

$$\alpha_j^e = \frac{exp(m_e^j)}{\sum_{k=1}^{M} exp(m_k^e)} \tag{10}$$

where W_e is the weight matrixes. α_j^e represents the importance of word in position j corresponding to the emotion expression. Usually, emotion expression is one word or short multiple-words which is not adapted to long term encoder. Therefore, we first feed emotion expression e into CNN to get the emotion expression $CNN(e)$. Then, we measure the importance of the word by making matrix multiplication between c' and $CNN(e)$. After that, the clause vector c_e and c_p are computed as a weighted sum of word annotations based on the weights α_j^e and α_j^p. For each clause, the final representation is the concatenation of c_e and c_p.

$$c_p = \sum(\alpha^p \cdot c) \tag{11}$$

$$c_e = \sum(\alpha^e \cdot c) \tag{12}$$

$$o = c_p \oplus c_e \tag{13}$$

$$y = softmax(W \cdot o) \tag{14}$$

The ECA classifier is trained by minimizing the cross entropy:

$$L = \sum_{(x,y)\in D'} \sum_{q \in Q} y^q log f^q(x;\theta) \tag{15}$$

where D' is the collection of training data and Q is the target category of sample. y^c is the target distribution. $f(x;\theta)$ is the predicted distribution of the model, θ is the parameter set.

Discriminator It is well acknowledged that the deeper fully connected network is, the better expressive ability it has. With the activiation layers and dropout layers are added, fully connected network not only has good ability for nonlinear mapping but also can prevent overfitting effectively. Therefore, considering the high prediction accuracy and the simple structure of fully connected network, we applied it to discriminator.

Output and Model Training For emotion cause classifier, we use the log likelihood of the correct labels as objective function:

$$\jmath(\Theta^m, \Theta^s) = \sum_{m=1}^{M} \sum_{i=1}^{N_m} logp(Y_i^{(m)} | c_i^{(m)}; \Theta^m, \Theta^s) \tag{16}$$

where Θ^m and Θ^s represent all parameters in private layers and shared layers. To guarantee that F can capture LI-TR features, discriminator D and shared feature extractor F need to game by adversarial training. For clause c, we construct discriminator D below,

$$p(\cdot|\Theta^d, \Theta^s) = softmax(W_d^T h_c^{(s)} + b_d) \tag{17}$$

Table 1. Cause Position of Each Emotion.

Distance	ECA-16	ECA-13-En-Train	ECA-13-En-Test
-3	1.60%	6.42%	6.23%
-2	5.53%	8.35%	7.63%
-1	32.90%	10.55%	9.38%
0	50.41%	12.84%	13.83%
1	5.22%	10.11%	8.18%
2	1.56%	7.76%	5.94%
3	0.48%	6.02%	4.68%
others	2.3%	37.95%	44.13%

Table 2. Distribution of emotion cause clauses and non-emotion cause clauses.

Distance	ECA-16	ECA-13-En-Train	ECA-13-En-Test
EC Clause	6.96%	14.17%	15.09%
NEC-Clause	93.04%	85.83%	84.90%

where $h_c^{(s)}$ is the output of shared feature extractor F. Θ^d represents the parameter W_d^T and b_d, Θ^s is the parameter on shared layer. For discriminator, the training objective is an adversarial gaming process, which consists of two components, i.e., optimizing the parameters of discriminator to identify the language of shared feature and optimizing the parameters of shared feature extractor to ensure that discriminator D cannot identify the language of shared feature. These two objective functions are as follows,

$$max_{\Theta^d} \jmath_{adv}^1(\Theta^s) = \sum_{m=1}^{M} \sum_{i=1}^{N_m} logp(m|c_i^{(m)}; \Theta^d, \Theta^s) \tag{18}$$

$$max_{\Theta^s} \jmath_{adv}^2(\Theta^d) = \sum_{m=1}^{M} \sum_{i=1}^{N_m} logH(p(m|c_i^{(m)}; \Theta^d, \Theta^s)) \tag{19}$$

where $H(p) = -\sum_i p_i logp_i$ and it is the entropy of distribution p. Finally, the objective function of ECA and adversarial training are incorporated to obtain the ultima objective function.

$$\jmath(\Theta; D) = \jmath_{seg}(\Theta^m, \Theta^s) + \jmath_{adv}^1(\Theta^d) + \lambda \jmath_{adv}^2(\Theta^s) \tag{20}$$

Here, λ is a weight hyper parameter to control the degree of each portion.

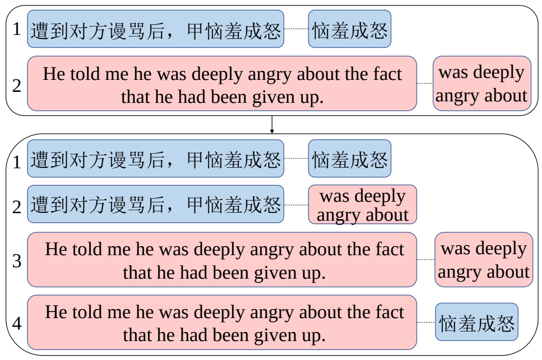

Fig. 3. The example of emotion cause extraction

4 Experiments

4.1 Data Sets

The proposed approach is evaluated on a Chinese emotion cause corpora[1] [3] (ECA-16) and NTCIR 2017 ECA-13 dataset[2] (ECA-13). The former one ECA-16 comprises 2,105 documents from SinaNews[3]. The latter one ECA-13 contains two components, i.e., 3,000 Chinese documents from social SinaNews and 3,000 English documents from novels. And for each sub-corpora of ECA-13, there are 2,500 documents for training and 500 documents for testing.

To observe the similarity and distinguish between ECA-13 and ECA-16, above all, we keep statistics on these datasets. The details are listed in Table 1 and Table 2. Table 1 shows that 88.53% emotion causes adjoin the emotion expressions, i.e., the distance is less one, however, only 33.5% in ECA-13-En-Train. Saliently, there are commonalities between these datasets. For example, emotion causes abut the emotion expressions. Apparently, position plays a very important role in ECA.

In this paper, we use 90% of the data from ECA-16 for Chinese feature extractor F_{cn} training and 10% for final testing. Besides, we employ ECA-13 English training dataset for English feature extractor F_{en} training. Specifically, we feed Chinese clauses and English clauses whose emotion categories are same to the former, English clauses and Chinese clauses whose emotion categories are same to the former, Chinese clauses, English clauses and all their position information into shared feature extractor F, as the example shown in Figure 3.

---

[1] http://hlt.hitsz.edu.cn/?page%20id=694
[2] http://hlt.hitsz.edu/?page_id=74
[3] https://news.sina.com.cn/society/

Table 3. Comparisons with existing methods.

Method	Precision	Recall	F1
RB	0.6747	0.4287	0.5243
CB	0.2672	0.7130	0.3887
RB+CB+ML	0.5921	0.5307	0.5597
MK_{SVM}	0.6673	0.6841	0.6756
ConvMS-Memnet	0.7067	0.6838	0.6955
ATCL-ECA	**0.7629**	0.6825	**0.7205**

4.2 Experimental Settings and Evaluation Metrics

In the experiments, we set the hidden units $h = 50$, the dimension of word embeddings $d = 100$ and the learning rate $lr = 0.002$. Specifically, the word embeddings are pretrained by word2vec from [18] and the 50-dimension position embeddings are randomly initialized with the uniform distribution U(-0.1,0.1). Dropout is set to 0.25 to overcome the overfitting in training process. And the batch size is set to 32 according to the best F value on the testing set. We evaluate the performance of emotion cause identification by the metrics used in [3], i.e. precision (P), recall (R), and F-measure (F), which is commonly accepted. If a proposed emotion clause covers an annotated answer, it is considered correct.

4.3 Comparisons of Different Methods

We compare ATCL-ECA with some traditional and advanced baselines as shown in Table 3.

- **RB** (Rule-based method): RB is proposed by [9], which is a rule-based approach to extract emotion cause.
- **CB** (Commonsense based method): The method is proposed by Russo [19], which uses the Chinese Emotion Cognition Lexicon as commonsense knowledge base [20].
- **RB+CB+ML** (Machine Learning): Gui [3] employed rule-based and commonsense based method to extract features and classify with machine learning algorithm.
- **MK**$_{SVM}$: Gui [3] used multi-kernel method trained with SVM to identify the emotion cause.
- **ConvMS-Memnet**: ConvMS-Memnet is proposed by Gui [4]. They regarded the task as a question answering problem, which is the current state-of-the-art method on emotion cause extraction.
- **ATCL-ECA**: ATCL-ECA is our proposed adversarial training based approach.

Table 3 shows the evaluation results, and we can observe the followings. Since rule-based methods face low coverage and poor universality, RB gives low recall.

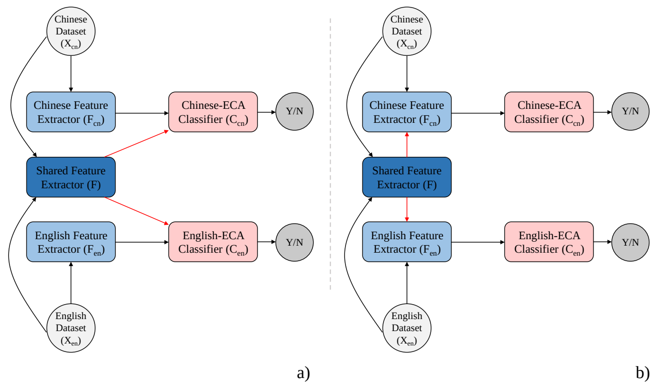

Fig. 4. a): The architecture of model-1; b): The architecture of model-2.

Table 4. The results of different architectures.

Method	Precision	Recall	F1
Model-1	0.6872	0.6733	0.6802
Model-2	0.6631	0.6530	0.6580
Model-3	**0.7629**	**0.6825**	**0.7205**

Nevertheless, commonsense based method achieves quite high recall but low precision. It is on account of there are almost all collocations between emotion expressions and emotion cause events when constructing the commonsense knowledge base, which ignores that the semantic information of the emotion expression is associated with contexts around it. RB+CB+ML verifies that RB and CB are complementary to improve the model performance. In addition, machine learning based method MK_{SVM} and deep learning based method ConvMS-Memnet outperform above approaches. Both of these two approaches consider the contextual information. Concretely, MK_{SVM} captures the structured information and lexical features. The latter ConvMS-Memnet can model the relations between emotion expressions and emotion cause clauses. The best F value is achieved by ATCL-ECA, which outperforms the state-of-the-art method ConvMS-Memnet by 2.5%. The results illustrate the performance of leveraging the cross-lingual semantic information. Meanwhile, it also verifies that cross-lingual information can bridge the gap between two languages effectively.

4.4 Comparisons of Different Architectures

We further study the performance of different architectures of cross-lingual emotion cause extraction. As shown in Figure 4 and Figure 1, the distinction of three models is that the input features fed into emotion cause extraction classifiers are diverse. **Model-1**: The concatenations of the features from shared feature extractor F and Chinese feature extractor F_{cn} are fed into Chinese ECA classifier

Table 5. The results of different sampling methods.

Sampling Method	Precision	Recall	F1
N-S	0.4749	0.7765	0.5894
U-S	0.6872	0.6654	0.6761
O-S	0.6945	0.6893	0.6919
O-S(batch 1:1)	**0.7629**	**0.6825**	**0.7205**

Table 6. The results of different attention hops.

Atention Hops	Precision	Recall	F1
Hop 1	0.6749	0.6625	0.6686
Hop 2	0.7049	0.6865	0.6856
Hop 3	**0.7122**	**0.7031**	**0.7076**
Hop 4	0.6739	0.6955	0.6845
Hop 5	0.6824	0.7024	0.6923
Hop 6	0.6705	0.6876	0.6789

C_{cn}. **Model-2**: The concatenations of the features from shared feature extractor F and embeddings of Chinese documents are fed into Chinese feature extractor F_{cn}. **Model-3**: Model-3 is the combination of Model-1 and Model-2. The concatenations of the features from F and embeddings of Chinese documents are fed into F_{cn}. Meanwhile, the concatenations of the features from F and F_{cn} are fed into Chinese ECA classifier C_{cn}.

The results listed in Table 4 are the comparisons of above three models. It is illustrated that the results of Model-3 outperforms others. Specifically, in Model-3, the concatenations of features from shared feature extractor F and private feature extractor F_{cn} and F_{en} play important roles in ECA classifiers. Moreover, for private feature extractors, the inputs are from pretrained embeddings of data and the outputs of shared feature extractor. Obviously, Model-3 could capture more cross-lingual semantic information for emotion cause extraction.

4.5 Effects of Sampling Methods

The different methods of sampling could reflect the strength of model learning ability. In this paper, the distribution of positive and negative samples in datasets is fairly imbalanced which is shown in Table 2, thus, the data requires sampling first. The results are listed in Figure 5. N-S represents no sampling, which gives low recall, precision and F value. Oversampling with batch 1:1 (O-S batch 1:1) achieves best performance compared with undersampling (U-S) and oversampling (O-S). Moreover, batch 1:1 adds not only the ratio of positive and negative samples to 1:1, but also the mandatory ratio to 1:1 on each batch. It is demonstrated that both sampling methods and the ratio of positive and negative samples in each batch play important roles to model optimizing process.

4.6 Effects of Different Attention Hops

It is well know that computational models using deep architecture with multiple layers could have better ability to learn data representations. In this section, we evaluate the influence of multiple hops in this task. We set the number of hops from 1 to 6. As shown in Table 6, the performance improves with the increasing number of hops when it is from 1 to 3. However, the performance decreases when the number of hops is larger than 3 because of the overfitting on this small dataset. Thus, we opted 3 hops in our final model since it gives the best performance.

5 Conclusion and Future Work

In this paper, we propose a new approach to identify the emotion cause corresponding to the emotion expression. The key property of this approach is the use of cross-lingual shared knowledge. The proposed model capture language-independent but associated with emotion cause extraction information by adversarial training. Instead of large-scale parallel corpora, our model achieves significantly better performance only on the small-scale Chinese corpora and English corpora compared to a number of competitive baselines. Meanwhile, the attention mechanism based on position information and emotion expression is designed to obtain the key parts of the clause. Experimental results verifies that our proposed approach outperforms a number of competitive baselines. In the future, we will construct English corpora from social news to shrink the disparity on cross-lingual corpora.

Acknowledgement

This work was supported by National Natural Science Foundation of China U1636103, 61632011, 61876053, Key Technologies Research and Development Program of Shenzhen JSGG20170817140856618, Shenzhen Foundational Research Funding 20170307150024907.

References

1. Lee, S.Y.M., Chen, Y., Huang, C., Li, S.: Detecting emotion causes with a linguistic rule-based approach. Computational Intelligence **29** (2013) 390–416
2. Chen, Y., Lee, S.Y.M., Li, S., Huang, C.: Emotion cause detection with linguistic constructions. In: COLING 2010, 23rd International Conference on Computational Linguistics, Proceedings of the Conference, 23-27 August 2010, Beijing, China. (2010) 179–187
3. Gui, L., Wu, D., Xu, R., Lu, Q., Zhou, Y.: Event-driven emotion cause extraction with corpus construction. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016. (2016) 1639–1649

4. Gui, L., Hu, J., He, Y., Xu, R., Lu, Q., Du, J.: A question answering approach to emotion cause extraction. CoRR **abs/1708.05482** (2017)
5. Banea, C., Mihalcea, R., Wiebe, J., Hassan, S.: Multilingual subjectivity analysis using machine translation. In: 2008 Conference on Empirical Methods in Natural Language Processing, EMNLP 2008, Proceedings of the Conference, 25-27 October 2008, Honolulu, Hawaii, USA, A meeting of SIGDAT, a Special Interest Group of the ACL. (2008) 127–135
6. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., Bengio, Y.: Generative adversarial nets. In: Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada. (2014) 2672–2680
7. Chen, X., Athiwaratkun, B., Sun, Y., Weinberger, K.Q., Cardie, C.: Adversarial deep averaging networks for cross-lingual sentiment classification. CoRR **abs/1606.01614** (2016)
8. Qinghong Gao, Jiannan Hu, R.X.L.G.Y.H.K.F.W.Q.L.: Overview of ntcir-13 eca task. In: Proceedings of the 13th NII Testbeds and Community for Information Access Research Conference on Evaluation of Information Access Technologies, NTCIR-13. (2017)
9. Lee, S.Y.M., Chen, Y., Huang, C.R.: A text-driven rule-based system for emotion cause detection. In: Naacl Hlt Workshop on Computational Approaches to Analysis Generation of Emotion in Text. (2010)
10. Ghazi, D., Inkpen, D., Szpakowicz, S.: Detecting emotion stimuli in emotion-bearing sentences. In: Computational Linguistics and Intelligent Text Processing - 16th International Conference, CICLing 2015, Cairo, Egypt, April 14-20, 2015, Proceedings, Part II. (2015) 152–165
11. Wan, X.: Co-training for cross-lingual sentiment classification. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. (2009) 235–243
12. Li, S., Wang, R., Liu, H., Huang, C.R.: Active learning for cross-lingual sentiment classification. Communications in Computer Information Science **400** (2013) 236–246
13. Zhou, X., Wan, X., Xiao, J.: Cross-language opinion target extraction in review texts. In: 12th IEEE International Conference on Data Mining, ICDM 2012, Brussels, Belgium, December 10-13, 2012. (2012) 1200–1205
14. Zhou, H., Chen, L., Shi, F., Huang, D.: Learning bilingual sentiment word embeddings for cross-language sentiment classification. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers. (2015) 430–440
15. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States. (2013) 3111–3119
16. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. CoRR **abs/1409.0473** (2014)

17. Fan, C., Gao, Q., Du, J., Gui, L., Xu, R., Wong, K.: Convolution-based memory network for aspect-based sentiment analysis. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018. (2018) 1161–1164
18. Zou, W.Y., Socher, R., Cer, D.M., Manning, C.D.: Bilingual word embeddings for phrase-based machine translation. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, EMNLP 2013, 18-21 October 2013, Grand Hyatt Seattle, Seattle, Washington, USA, A meeting of SIGDAT, a Special Interest Group of the ACL. (2013) 1393–1398
19. Russo, I., Caselli, T., Rubino, F., Boldrini, E., Martínez-Barco, P.: Emocause: An easy-adaptable approach to extract emotion cause contexts. In: Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, WASSA@ACL 2011, Portland, OR, USA, June 24, 2011. (2011) 153–160
20. Ruifeng, X.U., Zou, C., Zheng, Y., Jun, X.U., Gui, L., Liu, B., Wang, X.: A new emotion dictionary based on the distinguish of emotion expression and emotion cognition. Journal of Chinese Information Processing (2013)