Joint Learning for Emotion Classification and Emotion Cause Detection

Ying Chen[1], Wenjun Hou[1], Xiyao Cheng[1], Shoushan Li[2]
[1] College of Information and Electrical Engineering, China Agricultural University, China
[2] Natural Language Processing Lab, Soochow University, China
{chenying,houwenjun,chengxiyao}@cau.edu.cn
lishoushan@suda.edu.cn

Abstract

We present a neural network-based joint approach for emotion classification and emotion cause detection, which attempts to capture mutual benefits across the two sub-tasks of emotion analysis. Considering that emotion classification and emotion cause detection need different kinds of features (affective and event-based separately), we propose a joint encoder which uses a unified framework to extract features for both sub-tasks and a joint model trainer which simultaneously learns two models for the two sub-tasks separately. Our experiments on Chinese microblogs show that the joint approach is very promising.

1 Introduction

The analysis of emotions in texts is an important task in NLP. Traditional studies treat this task as a pipeline of two separated sub-tasks: emotion classification and emotion cause detection. The former identifies the category of an emotion and the latter detects the cause of an emotion. This separated framework makes each sub-task more flexible to deal with, but it neglects the relevance between the two sub-tasks. In this paper, we explore joint approaches which can capture mutual benefits across the relevant two sub-tasks. To the best of our knowledge, this work is the first attempt to incorporate both emotion classification and emotion cause detection into a unified framework.

Although emotion classification relies on affective features and emotion cause detection needs event-based features, we propose a joint encoder which uses a unified framework to extract features for both emotion classification instances and emotion cause detection instances. Then, we propose a joint model trainer which simultaneously learns two models for the two sub-tasks separately. The experiments on Chinese microblogs show that our joint approach can effectively learn models for both sub-tasks.

2 Our Approach

2.1 Corpus

In this paper, we use the human-labeled emotion corpus provided by Cheng *et al*. (2017) as our experimental data (namely Cheng emotion corpus). To better explain our work, we adopt twitter's terminology used in Cheng *et al*. (2017). Cheng emotion corpus can be considered as a collection of subtweets. For each emotion in a subtweet, all emotion keywords expressing the emotion are selected, and then the class and the cause of the emotion are annotated. The emotion categorization used in Huang et al. (2016) is adopted, which includes four basic emotions (i.e., joy, angry, sad and fearful) and three complex emotions (i.e., positive, neutral and negative). E.g. in the following example, the class of the emotion keyword ("😭") is *sad*, and the cause of the emotion is "only I was at home again".

Chinese : 兴冲冲勒跑回家~~发现又是我一个人再家。。😭 早知道就去蹭饭了

English Translation: I was very excited to run back home. I found that only I was at home again.😭 If I knew it earlier, I would have a meal for free.

Figure 1: An example of a subtweet

Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 646–651
Brussels, Belgium, October 31 - November 4, 2018.

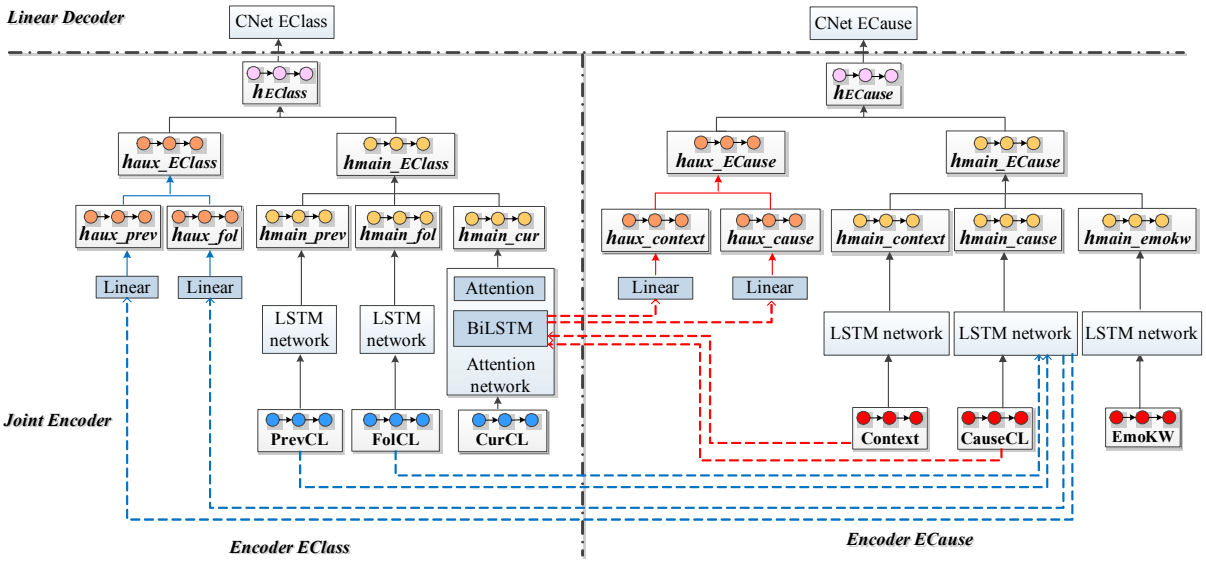

Figure 2: The framework of our joint approach

2.2 Problem Formulation

In this paper, both the emotion classification sub-task (namely *EClass*) and the emotion cause detection sub-task (namely *ECause*) are clause-level. Given an instance which is a clause in a subtweet, *EClass* assigns one of seven labels (i.e. six emotion classes and label '*non-emotion*' which indicates the absence of an emotion) to the instance. Notice, because of the extremely low percentage of emotion '*fearful*' (~0.6% in §3.1 Table 1), we ignore this emotion class in *EClass*. Given an instance which is a pair of <an emotion keyword, a clause in the subtweet>, *ECause* assigns a binary label to the instance to indicates the presence of a causal relation. Moreover, the clause-level *EClass* can effectively avoid the problem of multiple emotions (Li et al., 2015) because clauses are a kind of fine-grained texts.

Furthermore, the input text of an *EClass* instance contains three sequences of words: the previous clause (i.e. *PrevCL*), the current clause (i.e. *CurCL*), and the following clause (i.e. *FolCL*). The previous clause and the following clause provide contextual information for the current clause. The input text of an *ECause* instance also has three sequences of words: the emotion keyword (i.e. *EmoKW*), the current clause (i.e. *CauseCL*) and the context between *EmoKW* and *CauseCL*. The emotion keyword serves as an anchor, the current clause gives the description of an event which may cause the emotion, and the context provides complemental information for the event. Moreover, each word is represented with a vector from our word embedding model which is trained with word2vec[1] and the tweet corpus of Cheng *et al*. (2017).

2.3 The Joint Approach

As shown in Fig. 2, there are two parts in our joint approach which is based on neural networks: a joint encoder (the lower part) which extracts feature representations for both *EClass* instances and *ECause* instances, and a linear decoder (the upper part) which assigns labels to instances according to their representations.

Neural Networks

In the joint encoder, there are two neural networks (the attention network and the LSTM network), and each neural network is composed of several layers: bidirectional LSTM (i.e. BiLSTM) and attention. The BiLSTM layer focuses on the extraction of sequence features, and the attention layer focuses on the learning of word importance (weights). Because of the feature sparse problem in our small-scaled experimental data, the attention network often cannot effectively extract features to represent an event (see §3.2). Thus, in our joint encoder, we use the attention network to extract affective features (e.g. "😭" in Fig. 1) and the LSTM network to extract event-based features (e.g. "I found that only I was at home again" in Fig. 1).

[1] https://code.google.com/p/word2vec/

The attention network: we implement the attention network used in Felbo *et al*. (2017), which includes two layers: a BiLSTM layer which extracts a sequence feature for each input word, and an attention layer which represents the input sequence using weighted words.

The LSTM network: the network uses a BiLSTM layer to capture a sequence feature for each input word, and then uses the average of those features as the representation of the input sequence.

In the linear decoder, there are two classification networks (CNet *EClass* and CNet *ECause*) for *EClass* and *ECause* separately. Each classification network uses a linear layer to build a probabilistic classification model.

The Joint Encoder

As shown in Fig. 2, there are two sub-encoders in our joint encoder: Encoder *EClass* (the left part) which provides a representation for an *EClass* instance, and Encoder *ECause* (the right part) which extracts a representation for an *ECause* instance. Given an instance, one sub-encoder extracts a main representation (through the black lines in Fig.2) and the other sub-encoder provides an auxiliary representation (through the blue or red lines in Fig.2). Then, the concatenation of the two representations serves as the final representation for the instance (i.e. h_{EClass} or h_{ECause} in Fig.2). In order to deal with the case that a main representation may be overwhelmed by its corresponding auxiliary representation, linear layers are used to reduce the dimensions of auxiliary representations. Moreover, there are three sequences of words either in the input text of an *EClass* instance or in the input text of an *ECause* instance. In order to effectively use these input sequences, a multi-channel structure is chosen, which encodes the input sequences one by one.

Encoder *EClass*: given the three sequences of words in an *EClass* instance (*PrevCL*, *CurCL* and *FolCL*), the attention network is applied to *CurCL* to extract an affective representation, and the LSTM network is applied to *PrevCL* and *FolCL* separately to extract two event-based representations. Then, the concatenation of the three representations is used as the main representation (i.e. $h_{\text{main\_EClass}}$). Furthermore, in order to extract more contextual information, the LSTM network of Encoder *ECause* is applied to *PrevCL* and *FolCL* (through the blue lines in Fig. 2) to extract the auxiliary representation (i.e. $h_{\text{aux\_EClass}}$), which provides another event-based view for our emotion classification.

Encoder *ECause*: in order to separately deal with the three sequences of words (*EmoKW*, *CauseCL* and *Context*) in an *ECause* instance, the LSTM network is applied to each input sequence and then the concatenation of the three representations is used as the main representation (i.e. $h_{\text{main\_ECause}}$). Furthermore, for each input sequence (*CauseCL* or *Context*), the BiLSTM layer in the attention network is used to extract more event-based features (through the red lines in Fig. 2), and those features serve as an auxiliary representation (i.e. h_{aux_ECause}) which provides another event-based view for our emotion cause detection.

The Joint Model Trainer

During training, two models (*JMEClass* and *JMECause*) are learned simultaneously for the two sub-tasks (*EClass* and *ECause*) separately. Model *JMEClass* contains Encoder *EClass* and CNet *EClass*, and Model *JMECause* contains Encoder *ECause* and CNet *ECause*. Although each model uses auxiliary representations from the other model, but the learning of the model focuses on its own parameters. In other words, gradient calculation is disabled along the dashed lines in Fig. 2.

In each episode, the batch of input data is composed of two sets of instances: *EClass* sub-batch containing only *EClass* instances and *ECause* sub-batch containing only *ECause* instances. Given the batch of data, the parameters of each model are updated according its corresponding loss function. E.g., Model *JMEClass* uses only the *EClass* sub-batch, and its loss function is the mean squared errors of the instances in the sub-batch. In our joint model trainer, the two models are optimized using their own loss functions as pipeline model training does, but they use up-to-date auxiliary representations from each other to help optimization.

3 Experiments

3.1 Experimental Setup

In Cheng emotion corpus, there are ~3,000 sub-tweets, ~11,000 instances for *EClass*, and ~13,000 instances for *ECause*. Moreover, Table 1 lists the class distribution in Cheng emotion corpus for *EClass*. All experiments in this paper are

trained and tested by 5-fold cross-validation on Cheng emotion corpus, and all the results reported are the average ones of 5-fold cross-validation performances. We use the precision, recall and F-score as our evaluation metrics. However, because of the high percentage of label '*non-emotion*' in *EClass* (see Table 1) and label '*0*' in *ECause*, similar to previous work (Li et al. 2015; Felbo et al., 2017; Cheng *et al*., 2017; Gui *et al*., 2017), we report only the evaluation metrics of the six emotion classes for *EClass* and the evaluation metrics of label '*1*' for *ECause*.

Class	%	Class	%
Joy	11.3	Angry	3.5
Sad	2.6	Fearful*	0.6
Positive	8.2	Neutral	4.4
Negative	9.9	Non-emotion	59.5

Table 1: The class distribution in Cheng emotion corpus for *EClass*. (*: ignored)

During our joint training process, the dimension of the word embeddings is 20; the output dimension of the BiLSTM layer used in both the LSTM network and the attention network is 128; the output dimension of the linear network is 8; the batch size is 32.

The two models (*JMEClass* and *JMECause*) which are learned by our joint approach are compared with several pipeline models which are learned in a pipeline manner (i.e. either for *EClass* or for *ECause*) using one of the following state-of-the-art encoders.

- **ATT:** the attention network in Fig.2 .
- **LSTM:** the LSTM network in Fig.2.
- **ATT+LSTM:** an hybrid encoder for emotion classification, which applies *ATT* to *CurCL* and *LSTM* to *PrevCL* and *FolCL*.
- **ConvMSMemnet:** the encoder proposed by Gui et al. (2017) for emotion cause detection, which applies a convolutional multiple-slot deep memory network to *CauseCL*.

3.2 Method Analysis

Table 2 shows the performances of different emotion classification models, where "Sequence" lists the sequences of input words used by each model and each metric is the average performances of six emotion classes. Moreover, Table 3 lists the detailed performances of each emotion class in Model *JMEClass*.

Encoder	Sequence	Prec	Rec	F1
LSTM	*CurCL*	65.5	53.0	58.2
ATT	*CurCL*	67.6	56.5	61.0
	all	67.5	56.7	61.2
ATT+LSTM	all	67.0	57.8	61.7
JMEClass	all	***67.7***	***58.5***	***62.4***

Table 2: The performances of emotion classification models. (all: *PrevCL*, *CurCL* plus *FolCL*)

Class	Prec	Rec	F1
Joy	85.5	83.6	84.5
Angry	62.8	45.0	52.4
Sad	72.6	72.9	72.8
Positive	63.0	54.7	58.6
Neutral	62.5	41.2	49.7
Negative	59.9	53.9	56.7

Table 3: The performances of the six emotions in JMEClass

In Table 2, Model *ATT* + *CurCL* out-performs *LSTM* + *CurCL* by 2.8% in F-scores, where *ATT* is a state-of-the-art encoder for emotion classification (Felbo et al., 2017). The significant performance improvement means that *ATT* can effectively extract affective features in *CurCL*. In fact, the emotion classification on Chinese microblogs can rely much on emotion keywords occurring in *CurCL*. E.g. ~50% emotional instances in our experimental data contains emoticons (e.g. "😭" in Fig. 1) in *CurCL* and those emoticons themselves are strong emotion indicators. Secondly, when different kinds of contextual information are incorporated to Model *ATT* + *CurCL*, different performance improvements obtain (0.2% for *ATT* + *all* and 0.7% for *ATT*+*LSTM* in F-scores). This indicates that for the emotion classification, the event-based features extracted by *LSTM* are more helpful than the affective features extracted by *ATT*, because contexts often provide the cause event of an emotion. E.g. in Fig. 1, the previous clause of "😭" contains its cause "only I was at home again". Finally, taking the advantage of the event-based features extracted by *JMECause*, *JMEClass* out-performs the best pipeline model (*ATT+LSTM*) by 0.7% in F-scores. This shows that it is important for the emotion classification to have an encoder which can effectively extract event-based features from contexts.

In Table 3, the performance of a basic emotion (i.e., joy, angry or sad) is often better than the one of a complex emotion (i.e., positive, neutral or negative). However, in Table 1, the data size of a

basic emotion is often smaller than the one of a complex emotion. This indicates that difference in performance is likely linked to differences in the emotional contents of labels rather than differences in data sizes. E.g. the complex emotion 'negative' (i.e. a collection of complex emotions with negativity, such as 'hate', 'anxious', and so on) is more diverse than the basic emotion 'sad', and this diversity in emotional contents brings more challenges to the detection of this complex emotion. Furthermore, even if both 'sad' and 'angry' are basic emotions and have similar data sizes in our experimental data, it seems much easier to detect 'sad' instances than to detect 'angry' instances. This is maybe because 'angry' is caused by more various events and it is more difficult to capture and utilize those cause events. Thus, it is necessary for the emotion classification to have an encoder which can extract the event-based information of emotion cause from texts.

Encoder	Sequence	Prec	Rec	F1
ConvMS-Memnet	*CauseCL*	34.3	***77.5***	47.5
ATT	all	55.4	60.9	58.0
LSTM	all	***55.6***	61.3	58.3
JMECause	all	53.1	66.7	***59.1***

Table 4: The performances of emotion cause detection models. (all: *EmoKW*, *CauseCL* plus *Context*)

Table 4 shows the performances of different emotion cause detection models, where "Sequence" lists the sequences of input words used by each model. In Table 4, *JMECause* outperforms the best pipeline model (*LSTM*) by 0.8% in F-scores. The *LSTM* encoder is a state-of-the-art approach used for emotion cause detection (Cheng et al., 2017). Furthermore, the performance improvement of *JMECause* is from the significant increasing in recalls (5.4%). This indicates that more emotion causes are correctly detected when the event-based features extracted by Model *JMEClass* are incorporated. Moreover, among all models, the two models (*ATT* and *LSTM*) achieve relatively high precision and relatively low recall, and ConvMS-Memnet obtains the lowest precision and highest recall. This means that both *ATT* and *LSTM* suffer from the feature coverage problem because some useful features cannot be extracted through their encoders, and ConvMS-Memn suffers from the feature quality problem maybe because its encoder cannot handle the informal writing style used in Chinese microblogs.

4 Related Work

In recent years, intensive studies have explored supervised machine learning approaches using various types of features for different-level emotion classification, such as document level (Alm et al. 2005; Li et al. 2014; Huang et al. 2016), sentence level or short text level (Tokushisa et al. 2008; Bhowmick et al. 2009; Xu et al. 2012; Wen and Wan, 2014; Li et al. 2015; Felbo et al., 2017), and so on. Moreover, since both emotion and sentiment belong to affective feeling, some studies have explored the join learning of sentiment classification and emotion classification (Gao et al., 2013; Wang et al., 2015).

In the other hand, most of previous emotion cause detection studies is clause-based, which examine whether a clause around a given emotion keyword is a cause or not. Moreover, these studies (Chen *et al*., 2010; Xu *et al*., 2017; Ghazi *et al*., 2015; Gui *et al*., 2017; Cheng *et al*., 2017) focus on how to extract two kinds of features for supervised model learning: explicit expression patterns (e.g. "to cause", "for"), and implicit features which can reflect the causal relation.

5 Conclusion

In this paper, we focus on a joint learning approach to emotion classification and emotion cause detection on Chinese microblogs, and the experiments show such a joint approach is very promising.

Acknowledgments

This research work was partially supported by four the National Science Foundation of China (No. 61503386).

References

Cecilia Ovesdotter Alm, Dan Roth and Richard Sproat. 2005. Emotions from Text: Machine Learning for Text-based Emotion Prediction. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pages 579–586.

Plaban Kr. Bhowmick, Anupam Basu, Pabitra Mitra and Abhishek Prasad. 2009. Multi-label Text Clas-

sification Approach for Sentence Level News Emotion Analysis. In *Proceedings of the International Conference on Pattern Recognition and Machine Intelligence*, pages 261–266.

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

Xiyao Cheng, Ying Chen, Bixiao Cheng, Shoushan Li and Guodong Zhou. 2017. An Emotion Cause Corpus for Chinese Microblogs with Multiple-User Structures. In *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, Vol. 17, No. 1, Article 6.

Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1615–1625.

Wei Gao, Shoushan Li, Sophia Yat Mei Lee, Guodong Zhou and Chu-Ren Huang. 2013. Joint Learning on Sentiment and Emotion Classification. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 1505-1508.

Diman Ghazi, Diana Inkpen and Stan Szpakowicz. 2015. Detecting emotion stimuli in emotion-bearing sentences. In *Proceedings of the 16th International Conference on Computational Linguistics and Intelligent Text Processing*, 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 1593–1602.

Lei Huang, Shoushan Li and Guodong Zhou. 2016. Emotion Corpus Construction on Microblog Text. *The series Lecture Notes in Computer Science*, Volume 9332, pages 204-212.

Chengxin Li, Huimin Wu and Qin Jin. 2014. Emotion Classification of Chinese Miroblog Text via Fusion of Bow and Evector Feature Representations. In *Proceedings of the 3rd CCF Conference on Natural Language Processing & Chinese Computing*, pages 217-228.

Shoushan Li, Lei Huang, Rong Wang and Guodong Zhou. 2015. Sentence-level Emotion Classification with Label and Context Dependence. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 1045–1053.

Ryoko Tokuhisa, Kentaro Inui and Yuji Matsumoto. 2008. Emotion Classification Using Massive Examples Extracted from the Web. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 881-888.

Jun Xu, Ruifeng Xu, Qin Lu and Xiaolong Wang. 2012. Coarse-to-fine Sentence-level Emotion Classification based on the Intra-sentence Features and Sentential Context. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pages 2455-2458.

Ruifeng Xu, Jiannan Hu, Qin Lu, Dongyin Wu and Lin Gui. 2017. An Ensemble Approach for Emotion Cause Detection with Event Extraction and Multi-Kernel SVMs. *TSINGHUA SCIENCE AND TECHNOLOGY*, Volume 22, Number 6.

Rong Wang, Shoushan Li, Guodong Zhou and Hanxiao Shi. 2015. Joint Sentiment and Emotion Classification with Integer Linear Programming. In *Liu A., Ishikawa Y., Qian T., Nutanong S., Cheema M. (eds) Database Systems for Advanced Applications, Lecture Notes in Computer Science*, Vol 9052.

Shiyang Wen and Xiaojun Wan. 2014. Emotion Classification in Microblog Texts Using Class Sequential Rules. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 187-193.