Combining External Sentiment Knowledge for Emotion Cause Detection

Jiaxing Hu, Shumin Shi(✉), and Heyan Huang

School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China
{3220180705,bjssm,hhy63}@bit.edu.cn

Abstract. Emotion cause detection(ECD) that aims to extract the trigger event of a certain emotion explicitly expressed in text has become a hot topic in natural language processing. However, the performance of existing models all suffers from inadequate sentiment information fusion and the limited size of corpora. In this paper, we propose a novel model to combine external sentiment knowledge for ECD task, namely ExSenti-ECD, to try to solve these problems. First, in order to fully fuse sentiment information, we utilize a sentiment-specific embedding method to encode external sentiment knowledge contained in emotional text into word vectors. Meanwhile a new sentiment polarity corpus is merged from multiple corpora. Then, a pre-training method is adopted to mitigate the impact of the limitation of annotated data for ECD task instead of simply expanding samples. Furthermore, we apply attention mechanism to take emotional context into consideration based on the observation that the context around emotion keywords can provide emotion cause clues. Experimental results show that our model greatly outperforms the state-of-the-art baseline models.

Keywords: External sentiment knowledge · Pre-training · Emotional context · Emotion cause detection.

1 Introduction

In recent years, with the rapid development of modern social networking platforms such as Weibo, Weixin and Twitter, more and more people are getting accustomed to sharing their views and experiences with others on the Internet. In this situation, we have obtained a large amount of emotion and viewpoint information recorded in digital form. It has become a hot topic and challenge in the NLP field to mine emotion cause information from the massive data.

Emotion cause detection, as its name implies, is the process of discovering triggers or motivations that cause certain emotions. The task is totally different from the traditional tasks like emotion analysis and emotion prediction, where most of the current work is conducted. In many scenarios, we show more interest in the cause events of emotions rather than the certain emotion itself. For example, merchants prefer to know which attributes of their goods lead to customers' purchasing behavior. These attributes are generally included in the trigger events of emotion implicitly or explicitly.

There are several challenges of emotion cause detection that remain to be addressed. First, few corpora are publicly available for emotion cause detection. This is also the main reason for the lack of research on the task. The current mainstream machine learning and deep learning methods are all based on a large amount of tagged data to train their models. Second, the traditional word embedding methods do not capture enough semantic and sentiment information. It will cause significant downstream errors. For the lack of semantic information, for example, the word "bank" that appears in "bank deposit" and "river bank" apparently means different things. But they will have the same word embeddings if we apply methods like word2Vec to train their word vectors. Another instance will be given to demonstrate the absence of sentiment information. Words like "good" and "bad" that appear in the different sentences, that have the same syntactic structure but have completely opposite sentiment polarity, will be mapped to the close position in the vector space if we use traditional word embedding methods like word2Vec. This is undoubtedly a disaster for tasks related to sentiment analysis. For example, we would like to extract the cause events that express positive emotions corresponding to "feeling good", but because the distance between "good" and "bad" in vector space is very close, it may eventually extract the cause events that express negative emotions, which is certainly wrong. According to our observation of data sets, such samples account for a large proportion, so solving this problem will certainly bring a great improvement to the performance of our model.

To address the above-mentioned issues, a novel model called ExSenti-ECD is proposed in this paper. First, we try to encode external sentiment knowledge into our model in the phase of word embedding. Compared to general-purpose methods(word2Vec, Glove), the word vectors obtained by this way will fully fuse the sentiment information. This means the words like "good" and "bad" will be very far apart in vector space. Second, a well-known model called BERT in NLP field is introduced. The main reason why we consider using BERT can be summarized as two points: 1) the publicly available Chinese pre-trained model; 2)the deep bidirectional feature of the model. The first point will help solve the problem of the lack of corpora if we directly adopt the pre-trained model. Pre-training model has been proved to be an effective way in low-resources task. The second point shows the great advantage of BERT model in fusing context semantic information.

The main contributions of this paper can be summarized as follows: 1) A new sentiment polarity corpus is merged from several corpora to ensure enough external sentiment knowledge. 2) The sentiment-specific method is first utilized in ECD task to fully fuse sentiment information to reduce downstream error. 3) Pre-training method is introduced to mitigate the problem of poor model performance caused by the limited size of emotion cause corpora.

The remainder of this paper is organized as follows: The second section introduces the latest progress of related work. Details of our ExSenti-ECD model will be given in the third section. In the fourth section, we compare the experiment results of our model and other baselines to prove the validity of our proposed model. In the final section, we give a brief summary and possible future research directions.

2 Related Work

As a sub-problem in the field of NLP, the emotion cause detection task aims to mine deeper information of text. In the initial stage that the concept was proposed [4], the methods based on rules and common-sense knowledge are still the mainstream way to solve it [11, 1, 5, 10, 3]. With the further research, more and more methods have emerged. Gui et al. [6] use multi-kernel SVM based on the 7-tuple definition of event to extract emotion cause tuples. On the basis of FrameNet, Ghazi et al. [12] used CRF learners and a serialization model to identify emotional triggers in sentences containing emotions. Cheng et al. [13] proposed a multi-user structure to extract emotional cause events from Chinese microblog corpus using SVM classifier. The rise of deep learning methods also provides a new perspective on this issue. Deep memory network in QA is introduced to model the relationship between emotion keywords and candidate emotion cause clause [7]. Li et al. [8] further take into account of the context of the emotion clause to assist in selecting the right emotion cause clause. Chen et al. [9] propose a joint learning model of emotion classification and emotion cause detection. But very little effort has been devoted to construct corpora for the task [1, 5, 12]. And the corpus we used to train and evaluate our model is from Gui et al. [5].

In the field of NLP, most existing mainstream models use general-purpose pre-trained word embedding method [29, 28, 30] to get word vectors [19–21, 23–25]. However, in recent years, more and more attention has been paid to encode task-specific external knowledge into word embedding and pre-training method [27, 26]. Tang et al. [18] design and implement three neural networks to encode sentiment information into word vectors. Bespalov et al. [15] apply a n-gram model to integrate context information into word vectors. Labutov et al. [16] utilize existing resource to improve the performance of word embedding in the same vector space. Andrew et al. [17] proposed a hybrid supervised learning and unsupervised learning model, which fully fuses semantic and emotional information for word vector representation. In this paper, we propose a novel model to obtain as much of the sentiment information as it contains in the original text. In the end, the sentiment information obtained is reflected in the word embeddings of the text. The experimental results show that our model is superior to previous methods.

3 Our Model

3.1 Sentiment-Specific Embedding Method

Currently in NLP related tasks, general-purpose pre-trained word embedding methods like word2Vec, glove have become the preferred mainstream approaches. But there are still many shortcomings among these methods like that neither of them can fully integrate the emotional information contained in sentences. In recent years, in the field of emotional analysis, the method that encoding external emotional information into word vector representation to reduce downstream task errors has been proposed, and more and more attention has been paid to it.

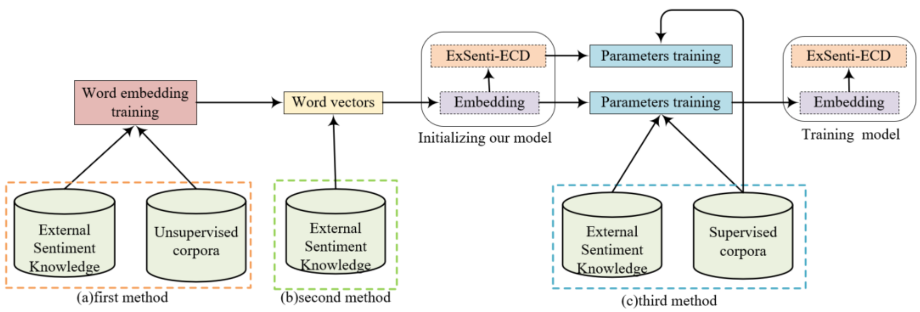

Fig. 1. Three methods for encoding external sentiment knowledge into word embedding

Fig. 1 illustrates three basic methods of encoding external sentiment knowledge into word vectors representation: (a) At the beginning of the training of word embedding model, external sentiment knowledge and unsupervised corpora are applied to the training model; (b) The pre-trained word embedding model is available. Then the pre-training model is fine-tuned based on external sentiment knowledge. The new model is used to initialize the word vectors; (c) External sentiment knowledge is combined with pre-trained word embedding in the process of joint parameters training. This enables the embedding module to be trained based on not only supervised training data, but also external sentiment knowledge.

The third method is used in our model proposed in this paper. And BERT is chosen to obtain word vectors after careful investigation and comparison. In order to integrate enough external sentiment knowledge into word vectors, a corpus containing more than 80,000 sentiment polarity samples is obtained by merging several sentiment polarity corpora from different fields like comments on social media or takeaway platform. This new corpus will be used to fine-tune the BERT pre-training model to encode external sentiment knowledge into the word embedding model, which is very useful for the tasks related to emotional analysis especially emotion cause detection.

3.2 Our ExSenti-ECD Model Based on BERT

BERT, which stands for Bidirectional Encoder Representations from Transformers, is a new language model proposed by Google in 2018 based on Transformers. The model swept 11 NLP task lists as soon as it was put forward and the result is amazing especially in fusing semantic information of text context. There are many ways to accomplish downstream NLP tasks by applying pre-training models, which can be roughly divided into feature-based methods(ELMO) and fine-tuning methods(ULMFit, OpenAI GPT). BERT implements both two methods and makes great improvements comparing with the existing methods. BERT uses masked language models to enable pre-trained deep bidirectional representations that can help fully fuse semantic information contained in text context. In addition, BERT doesn't require much task-specific design for certain tasks in different domains to achieve excellent generalization ability. This is why we choose it to be a necessary part of our model.

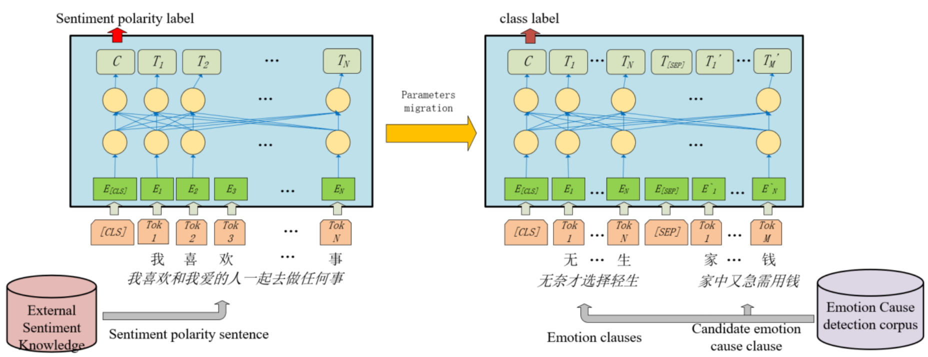

Fig. 2. High-level overview of our model.

Fig. 2 offers a high-level overview of our model. As the figure depicts, our ExSenti-ECD model can be roughly divided into two parts. In the first part, we try to pre-train a BERT model by applying the new sentiment polarity corpus we constructed from publicly available corpora. Through this process, we can endow the model higher ability to fuse sentiment information of text. This is also one of the core innovations proposed in this paper. Then the model parameters trained in the first part is used to initialize the second part of our ExSenti-ECD. In the second part, each input contains two parts: emotion clause and candidate emotional cause clause. And we use BERT to take emotional context information between them into consideration. Experimental results show that it can undoubtedly help improve the model performance.

The core concept of our ExSenti-ECD model is to introduce external sentiment knowledge so that word vectors of text can fully integrate the semantic and sentiment information to benefit emotion cause detection task. First of all, given the fact that the sizes of sentiment polarity corpora are generally small, we construct a new corpus that contains samples from several existing data sets. The samples contained in the new corpus only show positive or negative sentiment polarity. Then it is used to fine-tune the Chinese pre-trained BERT model. The word vectors of text are extracted by running the pre-trained model that has been well fine-tuned. Through the above several steps, we can get the word embedding matrices of emotion clauses and candidate emotion cause to represent them in the way that computers can process.

The corpus that used in our emotion cause detection task is constructed by Gui et al. [5]. Every sample is a triple tuple containing emotion clause, candidate emotion cause clause and a label that indicates whether the candidate emotion cause clause is a true one. Emotion clauses are the clauses that include one or more than one emotional keyword. Based on the observation that the context around emotion keywords can provide emotion cause clues because the emotion cause events are the stimuli of certain emotions, ExSenti-ECD model tries to apply attention mechanism to introduce the information of emotion clauses to assist in making decision. This method has been proved to be very effective [7, 8]. Multi-layer attention structures of BERT are directly adopted for the reason that the design of BERT is excellent enough and the available Chinese pre-trained model can mitigate problems caused by the limited size of data sets. The output of attention layers serves as the input of a softmax layer to get the final result. The reason for choosing such a simple composition of ExSenti-ECD model is still the limitation of data set. The model may not be able to converge if it is too complex.

4 Experiment Results

4.1 Data Sets

our ExSenti-ECD model requires two different corpora, one is sentiment polarity data set and the other is emotion cause data set. We get a relatively large sentiment polarity data set by aggregating several existing small data sets so that the new data set will contain enough sentiment information that can be used as external sentiment knowledge. The corpus only includes samples that show positive or negative sentiment polarity. In other words, no samples show neutral polarity. Considering that the new data set is used for encoding external sentiment knowledge and the neutral samples only account for a small proportion, we get rid of all neutral samples to eliminate their influence on experimental results. Table 1 shows the distribution of positive and negative samples in the corpus with a ratio of nearly one to one.

As shown in Table 2, our emotion cause corpus only has 2,105 documents that include 2,167 causes and 11,799 clauses. To solve this problem, we choose to work on model instead of simply expanding the data set because it is undoubtedly a time-consuming and laborious process to tag new data. And most documents have only one emotion cause event, so we can simplify the problem by assuming that all emotions correspond to only one cause event. We distinguish cause events

Table 1. Sample distribution in our new sentiment polarity corpus.

Polarity of Sentiment	Number
positive	40,051
negative	39,912

at the clause level, so for each emotion clause, only one emotion cause clause is the correct answer.

Table 2. Samples distribution of emotion cause corpus. Cause_1, Cause_2 and Cause_3 are emotions with 1, 2, and 3 emotion cause events respectively.

Items	Numbers
Documents	2,105
Causes	2,167
Clauses	11,799
Cause_1	2,046
Cause_2	56
Cause_3	3

4.2 Baselines

As illustrated in Table 3, we compare the experimental result of our ExSenti-ECD model with some classical methods to prove the validity of our model. These methods can be classified into three categories: (1) methods based on rules or common-sense knowledge; (2) machine learning methods; (3) deep neural network methods.

(1)Rule-based and common-sense based methods: **RB*** is a rule-based method which generalizes two sets of linguistic rules for emotion cause detection task [4]. **CB*** is a common-sense based method [11] that uses the Chinese Emotion Cognition Lexicon as the external knowledge base. **RB*** + **CB*** uses both rule-based and common-sense based methods to select right emotion cause event.

(2)Machine learning methods: **RB*** + **CB*** + **ML*** is a SVM classifier that features of rules and the Chinese Emotion Cognition Lexicon serve as training data set. **SVM*** is trained on unigrams, bigrams and trigrams features. **Multi-Kernel*** is also a SVM classifier but uses multi-kernel method [6].

(3)Deep neural network methods: **CNN*** is the classical convolutional neural network. **ConvsMS-Memnet*** is a new method proposed by Gui et al. [7] that use the multiple-slot deep memory network to select the correct emotion cause. **CANN*** use co-attention method to fuse the context information of emotion to help improve the task performance [8]. **ExSenti-ECD** is our model proposed in this paper.

From the experimental results that are showed in Table 3, we can draw some conclusions. First of all, for rule-based and common-sense based methods,

Table 3. Comparison of different word embedding methods. The results with superscript * are baselines from previous work.

Methods	P	R	F
RB*	**0.6747**	0.4287	0.5243
CB*	0.2672	**0.7130**	0.3887
RB* + CB*	0.5435	0.5307	**0.5370**
RB* + CB* + ML*	0.5921	0.5307	0.5597
SVM*	0.4200	0.4375	0.4285
Multi-Kernel*	**0.6588**	**0.6927**	**0.6752**
CNN*	0.5307	0.6427	0.5807
ConvsMS-Memnet*	0.7076	0.6838	0.6955
CANN*	0.7721	0.6891	0.7266
ExSenti-ECD	**0.8769**	**0.7062**	**0.7823**

RB+CB outperforms the two separate method RB and CB. It means the two methods can complement each other to some extent. And for machine learning methods, as we can see, the Multi-kernel method is obviously superior to the other two machine learning methods. The reason why it can perform so well is that the method take structured context information into consideration that other methods never do. For deep neural network methods, our ExSenti-ECD model greatly outperforms other neural network models in the precision rate meanwhile the recall rate also increased slightly. The improvement are undoubtedly owing to the integration of external sentiment knowledge and the adoption of pre-training methods and attention mechanism.

4.3 Comparison of Different Components in Our ExSenti-ECD Model

To verify the necessity of every part of our ExSenti-ECD model, we do another set of experiments that each of them makes a partial change to ExSenti-ECD model. As shown in Table 4, whether the external sentiment knowledge is not introduced or the context information of the emotion clauses is not added, the performance of the model will be worse especially in the recall rate. The huge increase in the recall rate indicates that our model can better predict the correct emotion cause events which is undoubtedly due to the addition of external sentiment knowledge and the context of emotion clauses.

4.4 Comparison of Different Word Embedding Methods

Although we have put forward improved strategies in other aspects, there is no doubt that encoding external sentiment information into word vector representation plays a crucial role. In order to verify it, we also experiment with some different word embedding methods that no any external knowledge is introduced. The experimental results are given in Table 5 and prove the effectiveness

Table 4. Effect of different components of ExSenti-ECD. Among them, No-Ex-Senti does not encode external sentiment knowledge into word embeddings, No-Attention means that we don't apply attention mechanism to introduce the context information of emotion clauses to help make decisions and ExSenti-ECD is our proposed model

Methods	P	R	F
No-Ex-Senti	0.85	0.5481	0.6665
No-Attention	0.8612	0.5654	0.6812
ExSenti-ECD	**0.8769**	**0.7062**	**0.7823**

of our proposed word embedding method. Random Initialization method randomly initializes word vectors for each word according to a certain probability distribution. Word2Vec, FastText and Glove are three currently popular word embedding methods. No-Ex-Senti means ExSenti-ECD model based on BERT but without encoding any external sentiment knowledge into it. No-Ex-Senti performs better than above four traditional word embedding methods and it demonstrates the great advance of the Chinese pre-trained BERT model in fusing context semantic information. Compared with No-Ex-Senti, the experimental result of ExSenti-ECD model has been further improved whether in precision, recall or F-measure. This proves that the introduction of external sentiment knowledge can indeed promote the effectiveness of the model because more sentiment information in the raw text can be captured for emotion cause detection.

Table 5. Comparison of different word embedding methods for emotion cause detection.

Methods	P	R	F
Random Initialization	0.8116	0.4030	0.5386
Word2Vec	0.7843	**0.482**	**0.5975**
FastText	0.7807	0.4776	0.5926
Glove	**0.8235**	0.4378	0.5717
No-Ex-Senti	0.85	0.5481	0.6665
ExSenti-ECD	**0.8769**	**0.7062**	**0.7823**

Fig. 3 was given to more clearly compare the performance of different word embedding methods under several distinct evaluation indicators including precision, recall and F-measure. Firstly, let's talk about the precision index. As the figure shows, all methods perform well on this indicator. Glove method is better than random initialization, word2Vec and fastText. Meanwhile our ExSenti-ECD model are the best among all traditional methods and No-Ex-Senti that contain no external sentiment knowledge. For the index of recall, only No-Ex-Senti and ExSenti-ECD methods exceed 50% and ExSenti-ECD performs better by at least 15% than other methods. Our ExSenti-ECD model also performs very well in F-measure, the most important indicator. The F-measure of our ExSenti-ECD is

78.23% while the highest of the other methods is 66.65%. The enormous increase in F-measure undoubtedly proves the validity of our proposed novel word embedding methods that encoding external sentiment knowledge into word vectors.

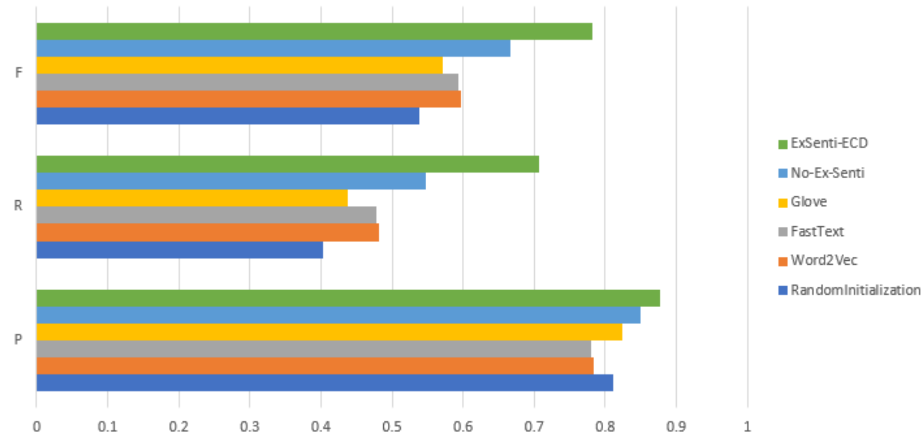

Fig. 3. Comparison of precision, recall and F-measure of different word embedding methods

5 Conclusion

In this paper, we propose a novel model called ExSenti-ECD to encode external sentiment knowledge into word vectors of text to benefit emotion cause detection task. The Chinese pre-trained BERT model is used for fusing sentiment and semantic knowledge and then get word vectors of text. The pre-training method can also solve the restriction of available data sets. The attention layers of BERT are applied to integrate extra context of emotion keywords to help select right emotion cause clause. In order to get enough external sentiment knowledge, we successfully merge several existing available sentiment polarity resources. The experimental results show that our ExSenti-ECD model greatly outperforms the existing baseline systems and other different word embedding methods.

The lack of tagged data is always an unavoidable problem for the training of deep neural network model. In future work, we will try to expand the data set for emotion cause detection rather than just improve model performance by applying new methods. On the other hand, currently most researchers focus on the identification of emotion cause of sentence expressing emotions explicitly while ignoring sentences expressing emotions implicitly. But examples of implicit expression of emotions meet the eye everywhere. Therefore, the concentration of our next work is to extract the emotion cause events of sentences that express emotion implicitly.

6 Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 61671064) and National Key Research&Development Program (Grant No. 2018YFC0831700).

References

1. Lee, S.Y.M., Chen, Y., Li, S., Huang, C.-R.: Emotion Cause Events: Corpus Construction and Analysis. In: LREC. ELRA, Valletta, Malta (2010)
2. Chen, Y., Lee, S.Y.M., Huang, C.-R.: A Cognitive-based Annotation System for Emotion Computing. In: Proceedings of the Third Linguistic Annotation Workshop (LAW III), pp. 1–9. Association for Computational Linguistics, Suntec, Singapore (2009)
3. Chen, Y., Lee, S.Y.M., Li, S., Zhou, G., Huang, C.-R.: Emotion Cause Detection with Linguistic Constructions. In: Proceedings of the 23rd International Conference on Computational Linguistics, pp. 179–187. Coling 2010 Organizing Committee, Beijing, China (2010)
4. Lee, S.Y.M., Chen, Y., Huang, C.-R.: 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, pp. 45–53. Association for Computational Linguistics, Los Angeles, CA (2010)
5. Gui, L., Yuan, L., Xu, R., Liu, B., Lu, Q., Zhou, Y.: Emotion Cause Detection with Linguistic Construction in Chinese Weibo Text. Natural Language Processing and Chinese Computing, 457–464 (2014)
6. Gui, L., Wu, D., Xu, R., Lu, Q., Zhou, Y.: Event-Driven Emotion Cause Extraction with Corpus Construction. In: EMNLP, pp. 1639–1649. ACL, Austin, Texas (2016)
7. Gui, L., Hu, H., He, Y., He, Y., Xu, R., Qin, L., Du, J.: A Question Answering Approach for Emotion Cause Extraction. In: EMNLP, pp. 1593–1602. ACL, Copenhagen, Denmark (2017)
8. Li, X., Song, K., Feng, S., Wang, D., Zhang, Y.: A Co-Attention Neural Network Model for Emotion Cause Analysis with Emotional Context Awareness. In: Conference on EMNLP, pp. 4752–4757. ACL, Brussels, Belgium (2018)
9. Chen, Y., Hou, W., Cheng, X., Li, S.: Joint Learning for Emotion Classification and Emotion Cause Detection. In: EMNLP, pp. 646–651. ACl, Brussels, Belgium (2018)
10. Neviarouskaya, A., Aono, M.: Extracting Causes of Emotions from Text. In: IJCNLP, pp. 932–936. AFNLP, Nagoya, Japan (2013)
11. Russo, I., Caselli, T., Rubino, F., Boldrini, E.: EMOCause: An Easy-adaptable Approach to Extract Emotion Cause Contexts. In: Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, pp. 153–160. Association for Computational Linguistics, Portland, Oregon (2011)
12. Ghazi, D., Inkpen, D., Szpakowicz, S.: Detecting Emotion Stimuli in Emotion-Bearing Sentences. In: International Conference on Intelligent Text Processing and Computational Linguistics, pp. 152–165. Springer International Publishing (2015)
13. Cheng, X., Chen, Y., Cheng, B., Li, S., Zhou, G.: An Emotion Cause Corpus for Chinese Microblogs with Multiple-User Structures. ACM Transactions on Asian and Low-Resource Language Information Processing, 1–19 (2017)

14. Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., Qin, B.: Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, pp. 1555–1565. ACL, Baltimore, Maryland (2014)
15. Bespalov, D., Bing, B., Qi, Y., Shokoufandeh, A.: Sentiment Classification Based on Supervised Latent n-gram Analysis. Acm International Conference on Information (2011)
16. Labutov, I., Lipson, H.: Re-embedding words. In: Proceedings of the 51st Annual Meeting of ACL, pp. 489–493. ACL, Sofia, Bulgaria (2013)
17. Maas, A., Daly, R., Pham, P., Huang, D., Ng, A., Potts, C.: Learning Word Vectors for Sentiment Analysis. In: 49th Annual Meeting of ACL, pp. 142–150. ACL, Portland, Oregon, USA (2011)
18. Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., Qin, B.: Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification. In: the 52nd Annual Meeting of ACL, pp. 1555–1565. ACL, Baltimore, Maryland (2014)
19. Luo, K., Deng, Z.-H., Yu, H., Wei, L.: JEAM: A Novel Model for Cross-Domain Sentiment Classification Based on Emotion Analysis. In: Proceedings of the 2015 Conference on EMNLP, pp. 2503–2508. ACL, Lisbon, Portugal (2015)
20. Zhu, S., Li, S., Chen, Y., Zhou, G.: Corpus Fusion for Emotion Classification. In: the 26th COLING, pp. 3287–3297. The COLING 2016 Organizing Committee, Osaka, Japan (2016)
21. Zhang, L., Wu, L., Li, S., Wang, Z., Zhou, G.: Cross-Lingual Emotion Classification with Auxiliary and Attention Neural Networks. NLPCC (2018)
22. Zhang, X., Zhang, C., Shi, H.: Ensemble of Binary Classification for the Emotion Detection in Code-Switching Text. Springer International Publishing (2018)
23. Mohtarami, M., Lan, M., Tan, C.: Probabilistic Sense Sentiment Similarity through Hidden Emotions. In: the 51st ACL, pp. 983–992. ACL, Sofia, Bulgaria (2013)
24. Gao, W., Li, S., Lee, S.Y.M., Zhou,G., Huang, C.R.: Joint learning on sentiment and emotion classification. In: the 22nd ACM, pp. 1505–1508. ACM, NY, USA (2013)
25. Ou, G., Chen, W., Wang, T., Wei, Z., Li, B., Yang, D., Wong, K.: Exploiting Community Emotion for Microblog Event Detection. In: EMNLP, pp. 1159–1168. ACL, Doha, Qatar (2014)
26. Devlin, J., Chang,Ming., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. (2018)
27. Ashish, V., Noam, S., Niki P., Jakob, U., Llion, J., Aidan, N., Lukasz, K., Illia, P.: Attention Is All You Need. Advances in Neural Information Processing, 6000–6010 (2017)
28. Mikolov, T., Sutskever, I., Chen, K.: Distributed Representations of Words and Phrases and their Compositionality. Advances in Neural Information Processing Systems, 3111–3119 (2013)
29. Mikolov, T., Corrado, G., Chen, K., Dean, J.: Efficient Estimation of Word Representations in Vector Space. 1–12 (2013)
30. Pennington, J., Socher, R., Manning, C.: Glove: Global Vectors for Word Representation. EMNLP, 1532–1543 (2014)