

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Computing and Information Systems

School of Computing and Information Systems

8-2021

Boundary detection with BERT for span-level emotion cause analysis

Xiangju LI

Wei GAO

Singapore Management University, weigao@smu.edu.sg

Shi FENG

Yifei ZHANG

Daling WANG

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research



Part of the [Theory and Algorithms Commons](#)

Citation

LI, Xiangju; GAO, Wei; FENG, Shi; ZHANG, Yifei; and WANG, Daling. Boundary detection with BERT for span-level emotion cause analysis. (2021). *Findings of the Association for Computational Linguistics*. 676-682. Available at: https://ink.library.smu.edu.sg/sis_research/6575

This Conference Proceeding Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylids@smu.edu.sg.

Boundary Detection with BERT for Span-level Emotion Cause Analysis

Xiangju Li¹, Wei Gao², Shi Feng¹, Yifei Zhang¹, Daling Wang¹

¹School of Computer Science and Engineering, Northeastern University, Shenyang, China

²School of Computing and Information Systems, Singapore Management University, Singapore
lixiangju100@163.com; weigao@smu.edu.sg; fengshi@cse.neu.edu.cn;
zhangyifei@cse.neu.edu.cn; wangdaling@cse.neu.edu.cn

Abstract

Emotion cause analysis (ECA) has been an emerging topic in natural language processing, which aims to identify the reasons behind a certain emotion expressed in the text. Most ECA methods intend to identify the clause which contains the cause of a given emotion, but such clause-level ECA (CECA) can be ambiguous and imprecise. In this paper, we aim at span-level ECA (SECA) by detecting the precise boundaries of text spans conveying accurate emotion causes from the given context. We formulate this task as sequence labeling and position identification problems and design two neural methods to solve them. Experiments on two benchmark ECA datasets show that the proposed methods substantially outperform the existing ECA models ¹.

1 Introduction

The task of emotion cause analysis (ECA) (Lee et al., 2010a), which is to extract the causes of an emotion expression from a given context, has gained increasing attention recently.

Most existing studies (Gui et al., 2016, 2017; Li et al., 2018, 2019; Xia et al., 2019) formulate ECA as a clause-level task (dubbed as CECA), which tries to extract the clauses that contain the emotion cause content. In CECA, a clause is typically a text segment separated by punctuation marks (e.g., ‘;’, ‘.’, ‘?’, ‘!’, etc.) in the given context. In the following example, the clause $[x_3]$ will be extracted as it contains the reason “the risk of infringement” that stimulates the emotion *afraid*.

Example 1. ² $[x_1]$ *The claim for damages caused by infringement has legal basis,* $[x_2]$ *and it is logical and reasonable.* $[x_3]$ *We can’t give up the good*

¹The code and datasets can be found at <https://github.com/xxxyyy2020/boundary-master>

²**Bold:** Gold emotion cause span; Underline: emotion expression. $[x_i]$ represents the i -th clause. The original report is in Chinese and translated into English.

*character of friendship and mutual assistance just because we are afraid of **the risk of infringement**.*
 $[x_4]$ *Give up food for being afraid of being choked.*
 $[x_5]$ *We should adhere to the traffic rules.*

However, determining the clause containing the stimulus is sub-optimal and inaccurate for ECA. In Example 1, while among the 5 clauses $[x_3]$ is the best, its main content “We can’t give up ... assistance” is not the cause of *afraid*. Such gap motivates a strong need of pinpointing more precise or finer-grained cause expressions which can convey the specific reasons of an emotion.

Some studies (Gao et al., 2015a,b; Neviarouskaya and Aono, 2013) try to extract emotion cause triples (a triple like (noun, verb, noun)) or emotion cause phrases. The extracted triples or phrases are usually not a complete emotion cause expression since the words in these triples or phrases are typically not continuous. Bi et al. (2020) proposed a new task for emotion and emotion cause span pair and classification task. Lee et al. (2010a) summarized seven groups of linguistic cues which could serve as indicators of cause events. Ghazi et al. (2015) built a Conditional Random Fields (CRF) learner to extract emotion cause spans with a set of manually engineered features. These two approaches are labor-intensive and prone to the sub-optimal design of features.

We study the Span-level ECA (SECA) based on state-of-the-art neural approach by focusing on detecting the boundaries of cause spans. We approach to the boundary detection in two different ways via sequence tagging and start/end position identification both in an end-to-end fashion. The emotion cause span is usually a sequence of consecutive tokens which convey exact reasons of the emotion and need to be inferred from the whole context. More specifically, we obtain the emotion

cause span by 1) mapping each token in the input context to a label indicating the ranges of spans; 2) pointing to the start and end positions of the span directly in the given context. The main contributions of our work are three-fold:

- We formulate SECA task as sequence tagging and position identification problems for boundary detection of emotion cause spans, and propose neural models to tackle this task.
- Different from previous ECA approaches, a pointer network is introduced to identify the start and end positions of emotion cause in our work. This is the first time to introduce the pointer network to ECA task.
- Experiments on two benchmark datasets show that our models substantially outperform the state-of-the-art ECA approaches at both span and clause levels.

2 Related Work

Lee et al. (2010b) first proposed a task on ECA and constructed a small-scale Chinese emotion cause corpus. Based on this corpus, Chen et al. (2010) proposed a rule based method for this task. Gui et al. (Gui et al., 2014) extended the rule-based method to Chinese microblog text. Gao et al. (Gao et al., 2015b,a) treated the emotion cause as a list of triples (a triple like (noun, verb, noun)). They structured a microblog post as a set of triples to determine which triple is the emotion cause. Shuntaro et al. (Yada et al., 2017) defined emotion cause as the nearest clause describing events of the cause of the emotion. Gui et al. (2016) constructed a public ECA dataset based on news corpus and proposed a multi-kernel-based method for CECA. All these methods rely on either manual rules or feature engineering.

Gui et al. (2017) first proposed a deep neural model for CECA. Following that, many CECA methods were developed with deep learning (Gui et al., 2017; Li et al., 2018; Yu et al., 2019; Ding et al., 2019; Li et al., 2019; Xia et al., 2019; Fan et al., 2019). Li et al. (2019; 2018) improved the performance of CECA by considering the contextual information of the emotion. Xia et al. (2019) proposed a hierarchical network for CECA based on Transformer encoder to capture different types of features. Fan et al. (2019) utilized the discourse information and clause position for the task. Xia

et al. (2019) and Chen et al. (2020a) proposed two variant tasks of ECA for extracting emotion and cause clause pairs. And then, many deep learning models (Wu et al., 2020; Chen et al., 2020b,c; Wei et al., 2020; Singh et al., 2021) were designed to tackle emotion cause pair extraction task.

Little work has been done for SECA. Lee et al. (Lee et al., 2010a) summarized the generalized rules for detecting the emotion causes of the emotion in Chinese. Based on manually crafted linguistic cues, Ghazi et al. (2015) built a CRF learner to identify emotion cause spans. However, these models are oversimplified and reliant on the ad-hoc features designed, thus are not generalized well.

3 Methodology

In this section, we first introduce two types definition of SECA. Then we describe our designed models.

3.1 Problem Formulation

Given a context S , which includes a sequence of n tokens $S = [w_1, w_2, \dots, w_n]$, the emotion expression $E = [e_1, e_2, \dots, e_m]$ and at least one emotion cause span, SECA task aims to detect the boundaries of emotion cause spans from S which stimulate the emotion expression.

Firstly, this task can be formulated as sequence labeling problem, of which the goal is to obtain the correct sequence of labels:

$$L = [l_1, l_2, \dots, l_n] \quad (1)$$

where $l_i \in \{B, I, O\}$, and B , I and O denote the beginning, inside of and outside of cause span, respectively, indicating the ranges of spans.

Secondly, this task also can be formulated as a position identification problem, where the goal is to get a list of correct start-end positions of emotion cause spans:

$$B = [(s_1, t_1), (s_2, t_2), \dots] \quad (2)$$

where s_i and t_i are the start and end token indices of the i -th cause span in the context, respectively.

3.2 Model Description

We use BERT (Devlin et al., 2019) as the backbone of our models due to it has strong contextualized representation ability. The context S and the emotion E are concatenated by forming a combined sequence as the input fed into BERT:

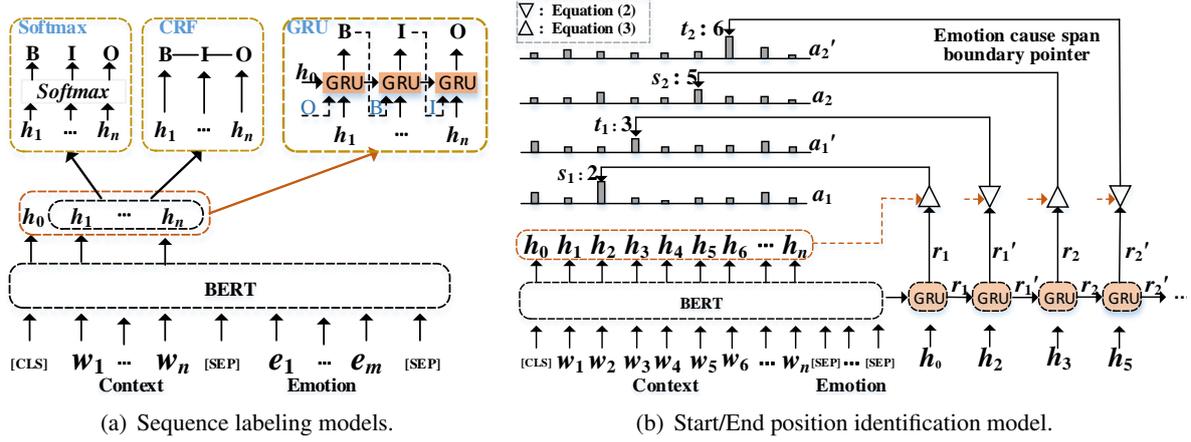


Figure 1: The architectures of the proposed boundary detection models for SECA.

$[\text{CLS}], w_1, \dots, w_n, [\text{SEP}], e_1, \dots, e_m, [\text{SEP}]$, where $[\text{CLS}]$ and $[\text{SEP}]$ are special tokens. Then, BERT receives the sequence and outputs the representation of each token in the combined sequence. The contextualized representations of each token $w_i \in S$ and the token $[\text{CLS}]$ can be given as:

$$h_i = \text{BERT}(w_i), \quad h_0 = \text{BERT}([\text{CLS}]) \quad (3)$$

where h_0 and $h_i \in \mathbb{R}^{d_b}$ and d_b is the vector dimension of the last layer of BERT. Based on the BERT encoding, we introduce the two types of span boundary detection models for SECA. An overview of our models are shown in Figure 1.

3.2.1 Sequence Labeling Models

We use three common ways to predict the labels of an input sequence. The first is the softmax function direct tag prediction which is straightforward to compute. The second is CRF, a well-known statistical graphical model which has demonstrated state-of-the-art accuracy on many sequence labeling tasks (Liu et al., 2014; Jin and Yu, 2021). The third is a variant of RNN (Goller and Kuchler, 1996) (e.g. GRU, LSTM) which generates tags sequentially by predicting the next label considering its previously predicted labels.

Softmax. Based on the token representation obtained in Equation (3), the probability distribution over the label set can be computed as:

$$p(l_i) = \text{softmax}(W_1 h_i + b_1) \quad (4)$$

where $W_1 \in \mathbb{R}^{d_b \times d_l}$ and $b_1 \in \mathbb{R}^{d_l}$ are learnable parameters, $p(l_i) \in \mathbb{R}^{d_l}$ is the label probability distribution of the i -th word and d_l is the size of label set.

CRF. CRF obtains the probability of a whole sequence:

$$p(L|S) = \frac{\exp(\text{score}(S, L))}{\sum_{L' \in \mathcal{L}} \exp(\text{score}(S, L'))} \quad (5)$$

where \mathcal{L} is the set of all candidate label sequences of the context. Here, we omit the detail of this classical prediction model (see Lafferty et al. (2001)).

GRU. The GRU version of RNN is effective and easy to train (Ma et al., 2019). At each predicting step, we update the hidden state h_i of token w_i using:

$$\tilde{h}_i = \text{GRU}(\tilde{h}_{i-1}, \ell(l_{i-1}) \oplus h_i) \quad (6)$$

where $\tilde{h}_i \in \mathbb{R}^{d_g}$ and d_g is the number of the GRU units, $\ell(l_{i-1}) \in \mathbb{R}^{d_\ell}$ is the embedding vector of previous label l_{i-1} and d_ℓ is the size of label embedding. Then, the probability distribution of l_i can be obtained by:

$$p(l_i) = \text{softmax}(W_2 \tilde{h}_i + b_2) \quad (7)$$

where $W_2 \in \mathbb{R}^{d_g \times d_l}$, $b_2 \in \mathbb{R}^{d_l}$ are training parameters.

3.2.2 Position Identification Model

Point network was first proposed by Vinyals et al. (2015), which is suitable for our positions identification task thanks to the fact that it is able to select positions from the input. Here we adopt it to generate the start and end positions of spans in turn. Because the number of cause spans is not fixed, we set a parameter \mathcal{C} to control how many spans to generate, and allow the two end points in (s_i, t_i) to take an integer value in $[0, n]$. For the k -th span (s_k, t_k) , we first get its start index s_k . An attention

mechanism is used to obtain the attention weight α_{kj} of the j -th token (for all $j \in [0, n]$) as follows:

$$a_{kj} = \frac{\exp(q_{kj})}{\sum_{j'=0}^n \exp(q_{kj'})} \quad (8)$$

$$q_{kj} = v^\top \tanh(h_j W_3 + r_k W_4)$$

where $W_3 \in \mathbb{R}^{d_b \times d_v}$, $W_4 \in \mathbb{R}^{d_r \times d_v}$ and $v \in \mathbb{R}^{d_v}$ are learnable parameters (d_v is the size of v), r_k is the hidden vector obtained by GRU:

$$r_k = \text{GRU}(h_{t_{k-1}}, r'_{k-1}) \quad (9)$$

where t_{k-1} is the index of end position of the $(k-1)$ -th span and $r'_{k-1} \in \mathbb{R}^{d_r}$ (d_r is the number of GRU units) is hidden vector obtained by GRU for predicting t_{k-1} (see below). Then, we set s_k as the highest attention weight from attention vector $a_k = [a_{k0}, \dots, a_{kn}]$.

To predict the end index t_k of the k -th span, the same attention mechanism is used to get the attention weight a'_{kj} :

$$a'_{kj} = \frac{\exp(\hat{q}_{kj})}{\sum_{j'=0}^n \exp(\hat{q}_{kj'})} \quad (10)$$

$$\hat{q}_{kj} = v^\top \tanh(h_j W_3 + r'_k W_4)$$

$$r'_k = \text{GRU}(h_{s_k}, r_k)$$

where W_3 , W_4 and v are learnable parameters. Similarly, the end index t_k can be identified by taking the maximum value in a'_k .

4 Experiments and Results

We evaluate the proposed methods on the Chinese ECA (**CHI**) dataset (Gui et al., 2016) and English ECA (**ENG**) dataset (Ghazi et al., 2015). On both datasets, cause spans are manually annotated and each context contains an emotion expression (or emotion category) and at least one cause span. The CHI dataset contains 2,105 instances and 2,147 cause spans. The ENG dataset contains 820 instances, and each contains one span only.

We use pre-trained *BERT-Base Chinese* and *BERT-Base Uncased* to encode the CHI and ENG datasets, respectively. We also choose the pre-trained span-level *SpanBERT-Base-cased* (n/a for Chinese) as the encoder on ENG dataset. We follow the default settings of BERT (Devlin et al., 2019) for fine-tuning. Adam (Kingma and Ba, 2015) optimizer is used with learning rate 1e-5. The epoch

Model	P^s	R^s	F_1^s
CHI			
Lee et al. (2010a)	0.197	0.161	0.177
Gui et al. (2017)	0.187	0.181	0.184
BERT _{base} +Softmax	0.483	0.574	0.525
BERT _{base} +GRU	0.481	0.567	0.520
BERT _{base} +CRF	0.564	0.570	0.566
BERT _{base} +Pointer	0.570	0.526	0.547
ENG			
Ghazi et al. (2015)	0.666	0.593	0.628
BERT _{base} +Softmax	0.838	0.876	0.856
BERT _{base} +GRU	0.883	0.868	0.875
BERT _{base} +CRF	0.866	0.890	0.878
BERT _{base} +Pointer	0.910	0.891	0.901
BERT _{span} +Softmax	0.830	0.891	0.859
BERT _{span} +GRU	0.893	0.879	0.886
BERT _{span} +CRF	0.858	0.883	0.870
BERT _{span} +Pointer	0.904	0.890	0.897

Table 1: SECA results based on span-level metrics.

is set to 10. The size of label embedding is 50. \mathcal{C}^3 is set to 3 on CHI and 1 on ENG which are the maximum number of cause spans in the respective training set. We follow the settings of previous works to split the datasets for train/dev/test (Ghazi et al., 2015; Gui et al., 2017). Hyper-parameters are tuned on the dev set.

We have BERT (**BERT_{base}**) and SpanBERT (**BERT_{span}**) as pre-trained models, and **Softmax**, **CRF**, **GRU** and pointer network (**Pointer**) as prediction models.

4.1 SECA Result

On CHI dataset, we compare our model with the rule based model (Lee et al., 2010a) and the word ECA model that outputs cause words (Gui et al., 2017), which is somewhat similar to our work. On ENG dataset, the CRF-based model in (Ghazi et al., 2015) is the only span-level baseline. For evaluation, we use span precision (P^s), recall (R^s) and F1 score (F_1^s) which are defined as

$$P^s = \frac{\# \text{ of correct cause spans}}{\# \text{ of predicted cause spans}}$$

$$R^s = \frac{\# \text{ of correct cause spans}}{\# \text{ of gold cause spans}} \quad (11)$$

$$F_1^s = \frac{2 \times P^s \times R^s}{P^s + R^s}$$

³If the number of emotion cause span(s) is less than \mathcal{C} , $(0, 0)$ is used for padding the correct list \mathcal{B} .

Table 1 shows that all our models strongly outperform the baselines, indicating our method is effective. More specifically, on CHI dataset our models BERT_{base}+Softmax, BERT_{base}+GRU, BERT_{base}+CRF and BERT_{base}+Pointer outperform the strong baseline model in Gui et al. (2017) by relative F_1^s improvements of 34.1%, 33.6%, 38.2% and 36.3%, respectively. The improvement is very significant with p -value less than 0.001 in t -test. With BERT encoder, Pointer performs the best on ENG and the second best on CHI, suggesting that pointer network is basically effective. Pointer is relatively worse than CRF on CHI because the number of spans it outputs is fixed as three while an instance on CHI may have less than three spans, rendering a small disadvantage. But generally, using pointer network for span boundary detection provides a strong alternative to classic CRF. It is a bit surprising that BERT_{span} does not show advantage over BERT_{base} here. We conjecture that given the small ENG dataset we cannot perform strong fine-tuning to make a big difference. In addition, the performance on ENG is much higher than on CHI because the context is generally much shorter making the task easier.

4.2 CECA Result

We can directly output clauses containing the predicted cause spans to compare with the rule based SECA model (2010a) and some strong CECA models (Gui et al., 2017; Li et al., 2018; Ding et al., 2019; Li et al., 2019; Xia et al., 2019; Fan et al., 2019; Hu et al., 2021). We only use CHI here since no clause labels are available and no previous work for CECA was done on English data for us to compare with. Following previous works (Gui et al., 2017; Xia et al., 2019), we use clause precision (P^c), recall (R^c), and F1 score (F_1^c) as evaluation metrics:

$$\begin{aligned}
 P^c &= \frac{\# \text{ of correct cause clauses}}{\# \text{ of predicted cause clauses}} \\
 R^c &= \frac{\# \text{ of correct cause clauses}}{\# \text{ of gold cause clauses}} \\
 F_1^c &= \frac{2 \times P^c \times R^c}{P^c + R^c}
 \end{aligned} \tag{12}$$

Table 2 shows that our proposed models outperform all the state-of-the-art CECA baselines. We attribute this to the fact that the BERT’s contextualized representation capacity and our SECA models are supervised by the finer-grained span-level annotations directly, which can effectively

Model	P^c	R^c	F_1^c
Lee et al. (2010a)	0.675	0.429	0.524
Gui et al. (2017)	0.708	0.684	0.700
Li et al. (2018)	0.772	0.689	0.727
Ding et al. (2019)	0.762	0.691	0.742
Li et al. (2019)	0.784	0.759	0.771
Xia et al. (2019)	0.770	0.766	0.768
Fan et al. (2019)	0.811	0.773	0.791
Hu et al. (2021)	0.786	0.757	0.771
BERT _{base} +Softmax	0.828	0.908	0.866
BERT _{base} +GRU	0.825	0.903	0.862
BERT _{base} +CRF	0.864	0.873	0.867
BERT _{base} +Pointer	0.834	0.844	0.838

Table 2: CECA results on CHI benchmark dataset.

guide model to learn more precise cause-related information. This advantage cannot be taken easily by the baseline approaches due to the nature of their clause-level supervision. Moreover, simply mapping the predicted spans to clauses for output makes Softmax become comparable with CRF and puts Pointer the last in terms of CECA performance. This is not surprising because an inaccurate span might still result in the right clause which just contains the predicted span. We also notice that rule based model performs worse than all the feature-learning models because that the manual rules hardly adopt to different datasets and feature-learning model can learn effective features according to the different datasets.

5 Conclusion

In this paper, we aim at span-level emotion cause analysis and propose neural sequence labeling and position identification models to detect the boundaries of emotion cause spans. The experiments conducted on two benchmark datasets of different languages demonstrate the effectiveness of the proposed approach, which achieves a new state-of-the-art performance on both span-level and clause-level ECA tasks.

Acknowledgments

The work was supported by National Key R&D Program of China (2018YFB1004700), National Natural Science Foundation of China (61772122,61872074,U1811261).

References

- Hongliang Bi and Pengyuan Liu. 2020. [ECSP: A new task for emotion-cause span-pair extraction and classification](#). *CoRR*, abs/2003.03507.
- Xinhong Chen, Qing Li, and Jianping Wang. 2020a. [Conditional causal relationships between emotions and causes in texts](#). In *In Proceedings of 2020 EMNLP*, pages 3111–3121.
- Xinhong Chen, Qing Li, and Jianping Wang. 2020b. [A unified sequence labeling model for emotion cause pair extraction](#). In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 208–218. International Committee on Computational Linguistics.
- Ying Chen, Wenjun Hou, Shoushan Li, Caicong Wu, and Xiaoqiang Zhang. 2020c. [End-to-end emotion-cause pair extraction with graph convolutional network](#). In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 198–207. International Committee on Computational Linguistics.
- Ying Chen, Sophia Yat Mei Lee, Shoushan Li, and Chu-Ren Huang. 2010. [Emotion cause detection with linguistic constructions](#). In *COLING 2010, 23rd International Conference on Computational Linguistics*, pages 179–187.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019*, pages 4171–4186.
- Zixiang Ding, Huihui He, Mengran Zhang, and Rui Xia. 2019. [From independent prediction to re-ordered prediction: Integrating relative position and global label information to emotion cause identification](#). In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019*, pages 6343–6350.
- Chuang Fan, Hongyu Yan, Jiachen Du, Lin Gui, Lidong Bing, Min Yang, Ruifeng Xu, and Ruibin Mao. 2019. [A knowledge regularized hierarchical approach for emotion cause analysis](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5618–5628.
- Kai Gao, Hua Xu, and Jiushuo Wang. 2015a. [Emotion cause detection for chinese micro-blogs based on ecocc model](#). In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 3–14.
- Kai Gao, Hua Xu, and Jiushuo Wang. 2015b. [A rule-based approach to emotion cause detection for chinese micro-blogs](#). *Expert Systems with Applications*, 42(9):4517–4528.
- Diman Ghazi, Diana Inkpen, and Stan Szpakowicz. 2015. [Detecting emotion stimuli in emotion-bearing sentences](#). In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 152–165.
- Christoph Goller and Andreas Kuchler. 1996. [Learning task-dependent distributed representations by backpropagation through structure](#). In *Proceedings of International Conference on Neural Networks (ICNN'96)*, volume 1, pages 347–352.
- 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. 2016. [Event-driven emotion cause extraction with corpus construction](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016*, pages 1639–1649.
- 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 - Third CCF Conference, NLPCC 2014*, volume 496 of *Communications in Computer and Information Science*, pages 457–464.
- Guimin Hu, Guangming Lu, and Yi Zhao. 2021. [FSS-GCN: A graph convolutional networks with fusion of semantic and structure for emotion cause analysis](#). *Knowl. Based Syst.*, 212:106584.
- Guozhe Jin and Zhezhou Yu. 2021. [A korean named entity recognition method using bi-lstm-crf and masked self-attention](#). *Comput. Speech Lang.*, 65:101134.
- Diederik P. Kingma and Jimmy Ba. 2015. [Adam: A method for stochastic optimization](#). In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. [Conditional random fields: Probabilistic models for segmenting and labeling sequence data](#). In *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001)*, pages 282–289.
- Sophia Yat Mei Lee, Ying Chen, and Chu-Ren Huang. 2010a. [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.
- Sophia Yat Mei Lee, Ying Chen, Shoushan Li, and Chu-Ren Huang. 2010b. [Emotion cause events: Corpus](#)

- construction and analysis. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010*.
- Xiangju Li, Shi Feng, Daling Wang, and Yifei Zhang. 2019. Context-aware emotion cause analysis with multi-attention-based neural network. *Knowledge-Based Systems*, 174:205–218.
- Xiangju Li, Kaisong Song, Shi Feng, Daling Wang, and Yifei Zhang. 2018. A co-attention neural network model for emotion cause analysis with emotional context awareness. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4752–4757.
- Yijia Liu, Yue Zhang, Wanxiang Che, Ting Liu, and Fan Wu. 2014. Domain adaptation for crf-based chinese word segmentation using free annotations. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 864–874. ACL.
- Dehong Ma, Sujian Li, Fangzhao Wu, Xing Xie, and Houfeng Wang. 2019. Exploring sequence-to-sequence learning in aspect term extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3538–3547.
- Alena Neviarouskaya and Masaki Aono. 2013. [Extracting causes of emotions from text](#). In *Sixth International Joint Conference on Natural Language Processing, IJCNLP 2013, Nagoya, Japan, October 14-18, 2013*, pages 932–936. Asian Federation of Natural Language Processing / ACL.
- Aaditya Singh, Shreeshail Hingane, Saim Wani, and Ashutosh Modi. 2021. [An end-to-end network for emotion-cause pair extraction](#). In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, WASSA@EACL 2021, Online, April 19, 2021*, pages 84–91. Association for Computational Linguistics.
- Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. 2015. [Pointer networks](#). In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 2692–2700.
- Penghui Wei, Jiahao Zhao, and Wenji Mao. 2020. [Effective inter-clause modeling for end-to-end emotion-cause pair extraction](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 3171–3181. Association for Computational Linguistics.
- Sixing Wu, Fang Chen, Fangzhao Wu, Yongfeng Huang, and Xing Li. 2020. [A multi-task learning neural network for emotion-cause pair extraction](#). In *ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, August 29 - September 8, 2020 - Including 10th Conference on Prestigious Applications of Artificial Intelligence (PAIS 2020)*, volume 325 of *Frontiers in Artificial Intelligence and Applications*, pages 2212–2219. IOS Press.
- Rui Xia and Zixiang Ding. 2019. [Emotion-cause pair extraction: A new task to emotion analysis in texts](#). In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 1003–1012. Association for Computational Linguistics.
- Rui Xia, Mengran Zhang, and Zixiang Ding. 2019. [RTHN: A rnn-transformer hierarchical network for emotion cause extraction](#). In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 5285–5291. ijcai.org.
- Shuntaro Yada, Kazushi Ikeda, Keiichiro Hoashi, and Kyo Kageura. 2017. [A bootstrap method for automatic rule acquisition on emotion cause extraction](#). In *2017 IEEE International Conference on Data Mining Workshops, ICDM Workshops 2017, New Orleans, LA, USA, November 18-21, 2017*, pages 414–421. IEEE Computer Society.
- Xinyi Yu, Wenge Rong, Zhuo Zhang, Yuanxin Ouyang, and Zhang Xiong. 2019. Multiple level hierarchical network-based clause selection for emotion cause extraction. *IEEE Access*, 7:9071–9079.