A Multi-Task MRC Framework for Chinese Emotion Cause and Experiencer Extraction

Haoda Qian[1,2](✉), Qiudan Li[2,3] and Zaichuan Tang[1,2]

[1] The State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
qiudan.li@ia.ac.cn

[2] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China
{qianhaoda19, tangzaichuan19}@mails.ucas.ac.cn

[3] Shenzhen Artificial Intelligence and Data Science Institute (Longhua)

Abstract. Extracting emotion causes and experiencers from text can help people better understand users' behavior patterns behind expressed emotions. Machine reading comprehension framework explicitly introduces a task-oriented query to boost the extraction task. In practice, how to learn a good task-oriented representation, accurately locate the boundary, and extract multiple causes and experiencers are the key technical challenges. To solve the above problems, this paper proposes BERT-based Machine Reading Comprehension Extraction Model with Multi-Task Learning (BERT-MRC-MTL). It first introduces query as prior knowledge and obtains text representation via BERT. Then, boundary-based and tag-based strategies are designed to select character to be extracted, so as to extract multiple causes or experiencers simultaneously. Finally, hierarchical multi-task learning structure with residual connection is adopted to combine the answer extraction strategies. We conduct experiments on two public Chinese emotion datasets, and the results demonstrate the efficacy of our proposed model.

Keywords: Emotion Cause and Experiencer Extraction · Machine Reading Comprehension · Multi-Task Learning

1 Introduction

Extracting emotion causes and experiencers from text is a fine-grained emotion analysis task. For instance, structual emotion phrases such as "陷入回忆" (caught up in memories)" and "白金跃 (Jinyue Bai)" extracted from Fig. 1 can provide a more comprehensive description of the "激动 (excited)" emotion, which answer the questions like "What is the cause of emotion?" and "Who feels such emotion?".

The machine reading comprehension based framework provides a task-specific extraction mechanism which guides the learning of extraction by explicitly introducing task-oriented query. This framework exerts easily generated queries and does not depend on particular category features. The key issue lies in how

当日，和中新网记者谈起建言献策的初衷，<情绪感受者>白跃金</情绪感受者><情绪原因>**陷入回忆**</情绪原因>，并略显<情绪词>**激动**</情绪词>。

On that day, when talking to a reporter from Zhongxin Net about the original intention of his advice and suggestions, <Experieincer>**Jinyue Bai**</Experiencer> was <Cause>**caught up in memories**</Cause> and slightly <Emotion keywords>**excited**</Emotion keywords>.

Fig. 1. An Example of Emotion Cause and Experiencer Extraction

to mine the relationship between query and content. Recently, pre-trained language models such as BERT [3] obtain good performance on machine reading comprehension (MRC) tasks for it contains rich linguistic knowledge [10] and can capture long-distance relations [2]. In this paper, we take BERT as our model backbone to acquire query-aware contextualized representation. In addition, traditional machine reading comprehension models can output only one answer given a query because it selects only the answer head and the answer tail with the maximum probability. However, in practice, there exists multiple emotion causes or experiencers for one emotion. To make up for this deficit, we propose boundary-based and tag-based answer extraction strategies to extract all causes or experiencers simultaneously. Moreover, as cause phrases have complex semantic structures and are variable in length, it is hard to accurately locate their boundaries. We design a multi-task learning structure to combine answer sequence labeling and answer boundary prediction into a unified framework, where hierarchical structure and residual connection are adopted to share the encoder parameters.

We conduct experiments on two public Chinese emotion datasets. The empirical results show that the BERT-based machine reading comprehension extraction model with multi-task learning achieves a boost over off-the-shelf sequence labeling models and named entity recognition models.

Contributions: Our contributions can be summarized as follows:

1) We propose a MRC-based framework that introduces task-oriented query to extract address emotion cause and experiencer extraction
2) Our model use contextualized Chinese character embeddings and does not suffer from segmentation errors or out of vocabulary (OOV) problems. The model uses easily generated query as guidance and fuses its prior knowledge into the content. The model also utilizes multi-task learning structure for more accurate extraction results.
3) We demonstrate the efficacy of the model on two public Chinese emotion datasets.

2 Related Work

Our work is related to structural emotion phrase extraction, machine reading comprehension framework and multi-task learning. In this section, we review the related works.

Extracting multiple emotion related structural information such as causes, experiencers and cues can enhance the interpretability of emotion analysis [17]. [11] adopted LSTM-CRF to extract causes, experiencers, targets and cues from English fiction corpus separately. [1] proposed an English news headline emotion corpus, and used BiLSTM-CRF to extract the emotion causes, experiencers, cues and targets. They formulate the extraction task as a common sequence labeling problem and didn't consider the relation between emotion roles and emotion expressed in the text, which limits its performance. [18] analyzed the emotion roles by masking or preserving specific emotion roles in the text when detecting emotions. Their results show that the cause and cue directly benefit emotion detection while experiencer is a confounder.

Most of the existing works focused on extracting cause and experiencer separately from English fiction or news headline corpus. Little work has been done on extracting emotion cause and experiencer from Chinese news content in a unified framework. News headlines are usually a summary of the news content and have higher information density. In contrast, the news content provides more details of an event. Hence, extracting cause and experiencer from news content facilitates understanding of the emotion behind the event from a deeper level and enables a multi-dimensional emotion analysis.

Machine reading comprehension provides a mechanism to extract answer spans from the content given a query. Such formulation has two main advantages. First, since query provides important prior information, it is highly task-oriented. Second, it does not rely much on manual features, and can easily apply to many scenarios. Recently, there has been a trend of converting NLP tasks [4] [12] [14] [15] into machine reading comprehension problem. [6] adopted memory slots to model context in- formation in emotion cause clause identification task. Different from these works, we focus on extracting structural information related to emotions.

3 Proposed Model

3.1 Problem Definition and Formulation

Given a news fragment $C = [c_1, c_2, ..., c_n]$, where n denotes the length of the news, our purpose is to find the structural information related to emotions using task-oriented query $Q = [q_1, q_2, ..., q_m]$, where m denotes the length of the query. We assign a label $y \in Y$ to c_i , where Y is a list of predefined phrase types (e.g., cause, experiencer, etc).

3.2 Model Architecture

The overview of our proposed model (BERT-MRC-MTL) is shown in Fig. 2. This model provides a unified framework that extracts structural emotional information. The model contains two modules: (1)A query-aware context encoder. The task-oriented query and news fragment are concatenated and put into BERT to get fully integrated. The encoder outputs the query-fused text representation.(2)A hierarchical multi-task learning answer extraction module with two-layer BiLSTM and residual connection. The contextualized token representation is input into the lower BiLSTM to tag the answer sequence, then the representation output by BERT and the output sequence from the lower BiLSTM are transfered together to the higher BiLSTM to predict the answer boundary. The final extraction results are decided by the higher BiLSTM.

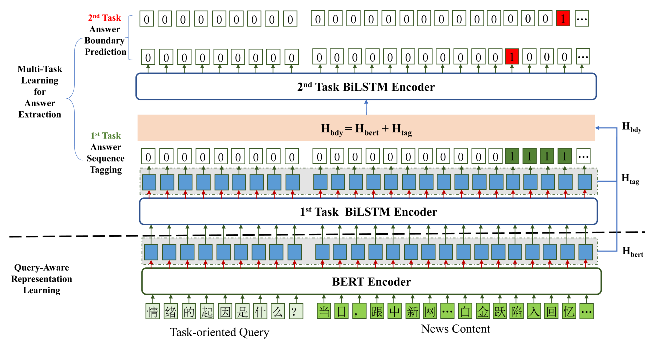

Fig. 2. The Overview of the Machine Reading Comprehension Extraction Framework with Multi-Task Learning

Query-Aware Context Encoder. Given the task-oriented query Q, we aim to extract the type y span from the news fragment. The key issue is to link and fuse the information from the query into the news fragment. Since BERT has brought a great boost to machine reading comprehension tasks, we employ it as our model backbone. To be in conformity with the input of BERT, we construct the input sequence $[[\mathrm{CLS}], Q, [\mathrm{SEP}], C]$. Then, the BERT receives the input and models the interactions between the query and the news using layer-by-layer Transformers, and outputs a context representation matrix $H_{bert} \in R^{n \times d}$, where d is the vector dimension of the last layer.

Answer Span Extraction. There may exist multiple causes or experiencers in one news fragment. Hence, we adopt two multi-span extraction strategies.

One strategy is a boundary-based one consisting of two binary token classifiers. One classifier predicts the probability of each token being a start, while the other predicts the probability of each token being an end:

$$L_{start} = \mathrm{Linear}(W_{start} H) \in R^{n \times 2} \tag{1}$$

$$L_{end} = \mathrm{Linear}(W_{end} H) \in R^{n \times 2} \tag{2}$$

Then, we apply argmax to L_{start} and L_{end} in the token level and get the starting and ending indices.

$$I_{start} = \{i \mid argmax(L^i_{start}) = 1, \quad i = 1, 2, \cdots, n\} \tag{3}$$

$$I_{end} = \{i \mid argmax(L^i_{end}) = 1, \quad i = 1, 2, \cdots, n\} \tag{4}$$

An issue to be considered is to generate pairs for the starting or ending indices. Since in our scenario there is no overlap between certain category, the Nearest Match Rule is applied to pair the start and end indices to compose answer. The total loss are losses for start and end indice predictions :

$$\mathcal{L}_{start} = \mathrm{BCE}(L_{start}, Y_{start}) \tag{5}$$

$$\mathcal{L}_{end} = \mathrm{BCE}(L_{end}, Y_{end}) \tag{6}$$

$$\mathcal{L}_{bdy} = \mathcal{L}_{start} + \mathcal{L}_{end} \tag{7}$$

where BCE denotes Binary Cross Entropy, Y_{start} and Y_{end} indicate the true start and end indices.

Another strategy is a tag-based one that casts answer extraction into sequence labeling and uses a binary classifier to output the probability distribution of tokens belonging to the answer:

$$L_{tag} = \mathrm{Linear}(W_{tag}H) \in R^{n\times 2} \tag{8}$$

$$I_{tag} = \{i \mid argmax(L^{i}_{tag}) = 1, \quad i = 1, 2, \cdots, n\} \tag{9}$$

This strategy can also extract a set of non-contiguous spans from the input text and does not need to handle the start/end matching issue. We gather those spans and get all the answers. The loss function is defined as follows:

$$\mathcal{L}_{tag} = \mathrm{BCE}(L_{tag}, Y_{tag}) \tag{10}$$

where Y_{tag} are the ground truth answer.

Multi-Task Learning Structure. Boundary-based extraction strategy selects the start and end indice of an answer. However, the loss will still be large even the predicted answer is close to the ground truth due to boundary prediction failure, and this strategy faces label imbalance problem. In contrast, tag-based answer extraction strategy labels the whole answer sequences, the loss function will be reduced with the predicted answers approximating the ground truth. Such strategy assigns the same weight to every token in the answer and is ought to put more emphasis on the boundary. It is intuitive that if we roughly locate the answer, it will be easier to predict its boundary. Inspired by the existing work of multi-task learning application in NLP tasks [5], we combine these two extraction strategies together to enhance the performance. Under the multi-task learning settings, the model is required to learn from both boundary-based and tag-based answer extraction tasks. The two subtasks share the BERT encoder, and each subtask has its corresponding classifier. Specifically, for the input in Fig. 2, to extract the cause phrase "陷入回忆 (fall into memory)", the answer tagging subtask will learn to tag the span "陷入回忆(fall into memory)", while the answer boundary prediction subtask will learn to predict the start indice "陷 (fall)" and the end indice "忆 (memory)".

We first come up with the parelle mode. This mode puts subtasks in a parallel manner. After the token representation is obtained, it is input into two BiLSTMs respectively for learning of answer extraction. One BiLSTM servers as the answer tagger and the other servers as the answer boundary predictor. This encourages the BERT to learn a hidden representation which benefits for both subtasks.

$$X_{tag} = X_{bdy} = H_{bert} \tag{11}$$

$$H_{tag} = \text{BILSTM}_{tag}(X_{tag}) \tag{12}$$

$$H_{bdy} = \text{BILSTM}_{bdy}(X_{bdy}) \tag{13}$$

Here *tag* and *bdy* represent answer tagging and boundary prediction.

Previous discussion indicates an order between the two subtasks. Thus, we train the two subtasks in hierachical order, where the answer tagger lies in the low level and the boundary predictor lies in the high-level. The text representation output by BERT is first input into the lower BiLSTM to tag answers:

$$X_{tag} = H \tag{14}$$

$$H_{tag} = \text{BILSTM}_{tag}(X_{tag}) \tag{15}$$

Then, the output hidden states of BiLSTM_{tag} are input into the higher BiLSTM to predict the answer boundary. In addition, to encourage information sharing between different layers and avoid error propagation, we add residual connection [8] between the BERT encoder and BiLSTM_{bdy}:

$$X_{bdy} = H_{bert} + H_{tag} \tag{16}$$

$$H_{bdy} = \text{BiLSTM}_{bdy}(X_{bdy}) \tag{17}$$

where BILSTM_{bdy} learns the answer boundary information. The final answer is decided by BILSTM_{bdy} using Eq.(1)(2)(3)(4).

Training Objective. Under multi-task learning settings, the overall training objective to be minimized can be summarized as follows:

$$\mathcal{L} = \lambda \mathcal{L}_{bdy} + (1 - \lambda) \mathcal{L}_{tag} \tag{18}$$

where λ is a hyper-parameter that controls the importance of each subtask. The two losses are jointly reduced in an end-to-end fashion, with parameters shared at the BERT encoder. In the experiment, we set λ to 0.5 based on the performance on the dev set.

4 Experimental Setup

4.1 Datasets

We conduct experiments on two public Chinese emotional datasets named HLTEmotionml and CEAC. HLEEmotionml[1] is a benchmark dataset for emotion cause analysis proposed by [7]. As there are only emotions and emotion causes annotated on this dataset, we ask two annotators to find experiencers for HLTEmotionml. The inter-annotator agreement of experiencer is 92%. CEAC[2] [16] labels emotion causes, experiencers, emotion keywords, actions on news corpus. For simplicity sake, we discard samples that lack cause or experiencer on both datasets. Table 1 shows the statistical information of the two datasets. During the experiment, we stochastically select 90% of the data for training and the remaining 10% for testing. For statistically credible results, we repeat the experiments 10 times and report the average result. For evaluation, we use the span-level micro precision, recall, and F1 score.

[1] HLTEmotionml is available at http://119.23.18.63/?page_id=694

[2] CEAC is available athttps://github.com/hupengyuan/EmotionAction_EmotionInference

Table 1. Dataset Stastical Information.

Dataset	HLTEmotionml	CEAC
Instances	2045	2915
Instances with multiple causes	57	515
Instances with multiple experiencers	42	6
Avg. text length	64.98	88.08
Avg. cause length	9.33	10.48
Avg. experiencer length	2.19	2.79

4.2 Compared Methods

Some widely-used sequence labeling models and named entity recognition models are introduced as baselines:

- LSTM-CRF [9]: It is a Chinese-character- based model using BiLSTM as encoder and CRF layer as decoder. Compared with word-based methods, this model does not suffer from segmentation errors.
- CAN-NER [19]: It is a convolutional attention network for Chinese named entity recognition using character embedding and segmentation information as features.
- BERT [3]: It uses BERT model as encoder with a token classification unit on top. This model treats extraction task as a sequence labeling problem.

We additionally propose a series of methods based on the MRC framework.

- BERT-MRC(Tag): It uses BERT to encode the query and the content, then adopts the tag-based span extraction strategy to extract answers.
- BERT-MRC(Bdy): Similar to BERT-MRC(Tag) but adopts boundary-based answer span extraction strategy.
- BERT-MRC-MTL(Parallel): A BERT-MRC model that combines tag-based span extraction strategy and boundary-based extraction strategy in a parallel way.
- BERT-MRC-MTL(Hierachy): A BERT-MRC model that combines tag-based span extraction strategy and boundary-based extraction strategy in a hierachical way.

4.3 Implementation Details

Under the MRC framework settings, the task-oriented query is instantiated as ”情绪的起因是什么？ (What is the cause of emotion?)” for cause extraction and ”情绪的感受者是谁？ (Who feels the emotion)” for experiencer extraction. We set the max sequence length for HLTEmotionml and CEAC to 120 and 160 respectively. For non-BERT models, the Chinese character embeddings are taken from[4] [13] . The character embedding size, hidden size of CNN and BiLSTM(Bi-GRU) are set to 300. The window size of CNN is set to 5, which is the same

[4] The Chinese character embedding is available at https://github.com/Embedding/Chinese-Word-Vectors

as [19]. Adam is used for optimization, with an initial learning rate of 0.001. For all BERT-based models, we use BEET-Base-Chinese[5] as backbone. The hidden dimension of BiLSTM in BERT-based models is 768. For BERT-Tagger and BERT-MRC(Tag), we train them with learning rate of 1e-5 for all parameters. For other models, we employ layer-wise learning rate, where BERT encoder has a learning rate of 1e-5 while the other layers have a learning rate of 1e-4. AdamW is employed for optimization. The hyper-parameters are selected based on the performance of the dev set. Note that we use Pytorch 1.6.0 and transformers 3.4.0 to implement our model.

5 Results and Discussions

The results are shown in Table 2.The models are categorized into non-BERT models and BERT-based models.

Table 2. F1-Score of the Extraction on the Two Datasets.

Dataset	HLTEmotionml			CEAC		
Type	Cause	Experiencer	All	Cause	Experiencer	All
Non-BERT Models						
LSTM-CRF	40.85	73.96	57.32	46.21	65.13	54.82
CAN-NER	44.75	76.1	60.53	47.61	68.29	56.94
BERT-based Models						
BERT-Tagger	54.52	87.68	70.52	61.94	82.35	70.79
BERT-MRC(Tag)	53.93	88.27	70.52	62.62	84.29	72.09
BERT-MRC(Bdy)	57.93	89.87	74.22	67.06	86.61	75.83
BERT-MRC-MTL(Parallel)	59.47	90.21	74.94	69.33	**87.03**	77.32
BERT-MRC-MTL(Hierachy)	**59.66**	**90.33**	**74.99**	**70.13**	86.95	**77.76**

5.1 Model Comparison

Evaluation on HLTEmotionml. For non-BERT models, CAN-NER achieves a F1 score of 60.53 and has a relative margin of 3.21 over LSTM-CRF. BERT-based models significantly outperform non-BERT models. BERT-MRC-MTL (Hierarchy) is the best among them with a F1 score of 74.99, which exceeds the best non-BERT model by over 14. BERT-MRC(Bdy) contains the target-oriented information and performs better than BERT-Tagger by 3.7 while BERT-MRC(Tag) has a comparable score to BERT-Tagger.
Evaluation on CEAC. CAN-NER is the best among non-BERT models and achieves a F1 score of 56.94, which is not satisfactory. BERT-MRC-MTL(Hierarchy)

[5] BERT-Base Chinese is available at https://huggingface.co/bert-basechinese/tree/main

obtains over 20 improvement in F1 over the best non-BERT method. BERT-MRC(Bdy) and BERT-MRC(Tag) beats BERT-Tagger by 5.04 and 3.7 respectively.

We observe that BERT-based models achieve larger improvements on CEAC than on HLTEmotionml. This phenomenon can be explained from two angles. On the one hand, CEAC is a larger dataset and the BERT encoder experiences more optimization steps.On the other hand, CEAC has a longer average text length and BERT can better captures long-distance dependencies than other encoders like LSTM.

5.2 Contributions of Modules

We conduct ablation tests to study on the contributions of different modules of the models.

Context Encoder. In this section, different context encoders are compared. LSTM-CRF uses BiLSTM as encoder to capture sequence-aware context information. CAN-NER consists of a CNN with local-attention to capture the implicit local context information and a GRU unit with global self-attention layer to capture information from adjacent characters and sentence context. With more local and context information, it outperforms LSTM-CRF by 3.21 on HLTEmotionml and 2.12 on CEAC. The lower F1 score on CEAC is due to its longer text length. BERT contains rich linguistic knowledge and can capture long-distance relations. Due to this reason, BERT-based models outperform non-BERT models significantly. The BERT encoder makes a larger improvement on CEAC due its longer text length.

Answer Extraction Strategy. We introduce task-oriented query and discuss two answer span extraction strategies. Compared with BERT-Tagger, BERT-MRC(Tag) is almost the same except for an additional task-oriented query. It achieves comparable performance on HLTEmotionml and a gain of 1.3 on CEAC. BERT-MRC(Bdy) outperforms BERT-Tagger by 3.7 on Emotional and 5.04 on CEAC. These results indicate that task-oriented query can enhance extraction, yet the chosen of span extraction strategies deserves consideration. The tag-based one assigns the same weight to all the answer tokens. In contrast, the boundary-based one emphasizes the importance of extracting the exact spans and can obtain more accurate answers. Based on this observation, we select the boundary-based strategy to compose the final answer.

Multi-Task Learning Structure. We perform analysis on multi-task learning settings. Directly applying boundary-based answer extraction strategy would suffer from label sparsity, while tag-based strategy can provide more category labels. In order to take advantage of both strategies, we co-train two answer extraction tasks.BERT-MRC-MTL(Parallel) settles the two tasks in the same level. This placement aims to share the parameters of context encoder.It achieves a relative margin of 0.72 and 1.49 against the best single task learning model. Intuitively, roughly locating the answer facilitates predicting its boundary. BERT-MRC-MTL(Hierachy) modifies the placement of the two subtasks to a hierarchical order. To avoid error propagation and encourage parameter sharing, it adds a residual connection to directly link the BERT and the higher-level BiLSTM. Compared with BERT-MRC(Parallel), BERT-MRC-MTL(Hierachy) allows the high-level task benefit from the low-level one and continues to make an inprovement on both datasets.

5.3 Case Study

Table 3. A Test Case on HLTEmotional.

Raw Text in Chinese and its English Translation		
S1.他愿意再去尝试。 (He is willing to try again.)		
S2.但是院方的建议是采用非血缘关系相匹配的骨髓或许成功率更高。 (But the hospital's advice is that using non-blood-matched marrow might have a higher success rate.)		
S3.如今康复出院的尹宾怡提起哥哥为了救她做5次骨髓移植的事， (When the recovered Yin Bingyi mentions the event that her brother had five hematopoietic stem cell transplantation to save her,)		
S4.感激的泪水就会夺眶而出。 (tears of gratitude will start from her eyes)		
Ground Truth		
Type	Cause	Experiencer
	哥哥为了救她做5次骨髓移植	尹宾怡
Prediction		
Model	Cause	Experiencer
LSTM-CRF	提起哥哥为了救她做5次骨髓移植的事(×)	尹宾怡(✓)
CAN-NER	提起哥哥为了救她做5次骨髓移植的事(×)	尹宾怡(✓)
BERT-Tagger	起(×)；哥哥为了救她做5次骨髓移植的事(×)	尹宾怡(✓)
BERT-MRC(Tag)	哥哥为了救她做5次骨髓移植的事(×)	尹宾怡(✓)
BERT-MRC(BDY)	哥哥为了救她做5次骨髓移植的事(×)	尹宾怡(✓)
BERT-MRC-MTL(Parallel)	哥哥为了救她做5次骨髓移植(✓)	尹宾怡(✓)
BERT-MRC-MTL(Hierachy)	哥哥为了救她做5次骨髓移植(✓)	尹宾怡(✓)

To show the predomination of our model vividly, we give an example from the test dataset on HLTEmotionml and compare the extraction results of different models in Table 3. First, it can be seen that all the models can extract experiencers correctly. This is mainly because experiencer is more like an entity-level role and is relatively easier to recognize. Second, when it comes to cause extraction, the results vary for different models. Non-BERT models can roughly locate the cause phrase but extract a much longer span than the gold annotation. This is because these models can only catch surface information and short-distance dependencies. BERT carries rich semantic knowledge and can capture long-distance dependencies, thus narrowing down the span by a little and closer to the gold annotation. BERT-MRC models utilizes query that contains prior knowledge query-aware representation and approximate the gold cause.

By multi-task learning, BERT-MRC-MTL models further make an advance and accurately predicts the cause boundary.

6 Conclusion

We present a novel extraction framework to obtain emotion cause and experiencer from news corpus, which introduces task-oriented query and deeply integrate it with content through BERT, and adopts multi-task learning for more accurate extraction spans. Experiments on two public Chinese emotion datasets show that BERT-MRC-MTL(Hierachy) outperforms other models. In future work, we will investigate in emotion cause and experiencer extraction in more complex scenarios such as Weibo, which contains dialogue structures and rich user interactions.

7 Acknowledgement

This work was supported in part by the National Key R&D Program of China under Grant No. 2016QY02D0305, the National Natural Science Foundation of China under Grant No. 61671450, 71621002, 71702181, the Key Research Program of the Chinese Academy of Sciences under Grant No. ZDRW-XH2017-3. This work was partially supported by the National Key Research and Development Program of China under Grant No. 2020AAA0103405, the National Natural Science Foundation of China under Grant No. 62071467, 71902179, 71621002.

References

1. Bostan, L.A.M., Kim, E., Klinger, R.: GoodNewsEveryone: A corpus of news headlines annotated with emotions, semantic roles, and reader perception. In: LREC (2020)
2. Clark, K., Khandelwal, U., Levy, O., Manning, C.D.: What does bert look at? an analysis of bert's attention. In: Proceedings of the 2019 ACL Workshop BlackboxNLP (2019)
3. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT (2019)
4. FitzGerald, N., Michael, J., He, L., Zettlemoyer, L.: Large-scale qa-srl parsing. ArXiv **abs/1805.05377** (2018)
5. Gong, Y., Luo, X., Zhu, Y., Ou, W., Li, Z., Zhu, M., Zhu, K.Q., Duan, L., Chen, X.: Deep cascade multi-task learning for slot filling in online shopping assistant. In: AAAI (2019)
6. Gui, L., Hu, J., He, Y., Xu, R., Lu, Q., Du, J.: A question answering approach for emotion cause extraction. In: EMNLP (2017)
7. Gui, L., Wu, D., Xu, R., Lu, Q., Zhou, Y.: Event-driven emotion cause extraction with corpus construction. In: EMNLP (2016)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 770–778 (2016)

9. Huang, Z., Xu, W., Yu, K.: Bidirectional lstm-crf models for sequence tagging. ArXiv **abs/1508.01991** (2015)
10. Jawahar, G., Sagot, B., Seddah, D.: What does bert learn about the structure of language? In: ACL (2019)
11. Kim, E., Klinger, R.: Who feels what and why? annotation of a literature corpus with semantic roles of emotions. In: COLING (2018)
12. Levy, O., Seo, M., Choi, E., Zettlemoyer, L.: Zero-shot relation extraction via reading comprehension. In: CoNLL (2017)
13. Li, S., Zhao, Z., Hu, R., Li, W., Liu, T., Du, X.: Analogical reasoning on chinese morphological and semantic relations. In: ACL (2018)
14. Li, X., Feng, J., Meng, Y., Han, Q., Wu, F., Li, J.: A unified mrc framework for named entity recognition. In: ACL (2020)
15. Liu, J., Chen, Y., Liu, K., Bi, W., Liu, X.: Event extraction as machine reading comprehension. In: EMNLP (2020)
16. Liu, P., Du, C., Zhao, S., Zhu, C.: Emotion action detection and emotion inference: the task and dataset. ArXiv **abs/1903.06901** (2019)
17. Oberländer, L., Klinger, R.: Token sequence labeling vs. clause classification for english emotion stimulus detection. ArXiv **abs/2010.07557** (2020)
18. Oberländer, L., Reich, K., Klinger, R.: Experiencers, stimuli, or targets: Which semantic roles enable machine learning to infer the emotions? ArXiv **abs/2011.01599** (2020)
19. Zhu, Y., Wang, G., Karlsson, B.F.: Can-ner: Convolutional attention network for chinese named entity recognition. In: NAACL-HLT (2019)