# Subject Recognition in Chinese Sentences for Chatbots

Fangyuan Li[1], Huanhuan Wei[1], Qiangda Hao[1], Ruihong Zeng[1], Hao Shao[1⋆], and Wenliang Chen[2]

[1] Gowild Robotics Co., Ltd, China
[2] School of Computer Science and Technology, Soochow University, China
{fyli.winnie, haoqiangda3, petertsengruihon, shaohao820}@gmail.com
weihuanhuan5@163.com, wlchen@suda.edu.cn

**Abstract.** Subject[3] recognition plays a significant role in the conversation with a Chatbot. The misclassification of the subject of a sentence leads to the misjudgment of the intention recognition. In this paper, we build a new dataset for subject recognition and propose several systems based on pre-trained language models. We first design annotation guidelines for human-chatbot conversational data, and hire annotators to build a new dataset according to the guidelines. Then, classification methods based on deep neural network are proposed. Finally, extensive experiments are conducted to testify the performance of different algorithms. The results show that our method achieves 88.5% $F_1$ in the task of subject recognition. We also compare our systems with three other Chatbot systems and find ours perform the best.

**Keywords:** Chatbot · Subject recognition · Deep neural network.

## 1 Introduction

As an entry-level product in the era of artificial intelligence, chatbots have attracted wide attention from both academic and industry. In a chatbot framework, intention recognition significantly influence the conversational performance. Nevertheless, a large number of chatbots performs worse in intention recognition, especially in the recognition of the subject of a sentence. As shown in Figure 1, although the chatbot successfully identifies specific topics including "singing" and "constellation", it fails to distinguish characters such as "you" and "me" around them. The underlying reason is that, situations such as missing subject, unknown reference and sentence ambiguity often occur in spoken Chinese. Consequently, it is more difficult to perform semantic analysis than in written Chinese.

To our best knowledge, there exists no effective resolution to the problem of character recognition in chatbots. A number of research works has been devoted to identifying the grammatical subjects [7, 11, 13], the agents [1, 10], the emotion

⋆ Corresponding author

[3] In this paper, subject means "主体/zhu ti" in Chinese, while we use "grammatical subject" to denote traditional "主语/zhu yu" in Chinese.

**User:** So boring. Can you sing a song for me?
**Chatbot:** OK. Twinkle twinkle little star, how I wonder what you are…
**User:** Great, let me sing a song for you.
**Chatbot:** OK. Twinkle twinkle little star, how I wonder what you are…
**User:** ……
**User:** By the way, my constellation is Libra.
**Chatbot:** You don't care much about me. I'm Aries.

**Fig. 1.** Example of subject error in chatbot

or opinion holders [4, 16] and targets extraction [12, 15]. However, none of these achievements can be directly used to identify core characters in a conversational context. In this paper, the term "subject" (主体/zhu ti) is defined as the characters around events, opinions and descriptions. It is different from the traditional "grammatical subject" (主语/zhu yu) in three aspects:

**1) The domains and scopes are different.** The grammatical subject of a sentence can be composed of nouns, pronouns, noun phrases, adjectives, verbs, predicate phrases and subject-predicate phrases. The subject in this paper pays attention to a specific person around the core content in a sentence. For example, in the sentence "learning to drive is difficult", the grammatical subject is the verb-object phrase "learning to drive". However, there is no specific character which can be denoted as a subject.

**2) The designations are different.** The grammatical subject is usually the declarative object of a sentence, while the subject is the specific person to whom the declarative object belongs. For example, in the sentence "my dream is beautiful", the grammatical subject is "dream", while the subject is "I".

**3) The numbers are different.** There may be multiple grammatical subjects in a sentence, but generally there is only one subject. In simple sentences, the grammatical subject is unique, but there may or may not exist a subject. In complex sentences, grammatical subjects may appear in both main clauses and subordinate clauses, however, only one character can be regarded as the subject. E.g., "I" and "you" are both grammatical subjects of "I think you are beautiful", while the subject is "you" around the core contents "beautiful".

In real applications, subject recognition is difficult due to incomplete semantics and unclear references, especially without context information. For example, "Cheer up mom loves you" (加油妈妈爱你), it is difficult to distinguish the real subject between "user's mom" or "user". At present, there is no widely acknowledged definition and annotation guidelines of subject recognition. To our best knowledge, there is not an annotated dataset available for this task. There are three main contributions of this paper:

- The first annotation guidelines of subject recognition are proposed.
- A dataset is constructed manually according to the annotation guidelines which can be downloaded publicly.
- Neural network based subject recognition methods are constructed, extensive experiments are conducted to verify the effectiveness from both the model evaluation aspect and the manual evaluation aspect.

## 2 Related Work

Subject recognition in the field of chatbots has certain relevance to many tasks in Natural Language Processing(NLP), such as Summarization, Sentiment Analysis and Semantic Role Labeling(SRL). Sentence-level summarization aims at extracting the Subject-Verb-Object(SVO) tuple in a sentence. Most of them focus on extracting the SVO based on syntactic analysis [7, 11, 13]. Sentiment analysis is the field of analyzing people's opinions, sentiments and emotions. Previous studies have been mainly dedicated to opinion mining and sentiment classfication [12, 17], while less attention is paid to the extraction of opinion holders and targets. Opinion holders extraction is usually based on linguistic rules which are constructed by named entity recognition and syntactic analysis features [4, 16]. Opinion targets are related to opinion words. Plenty of researchers extract them after recognizing the opinion words [12, 15]. Semantic Role Labeling studies the relationship between predicates and other components in sentences. The conventional approach to SRL [1, 10] is to construct a syntax tree and prune it with heuristic rules. Then classifiers are used to identify and classify potential arguments. Finally, global semantic roles are obtained by inference.

Subject recognition has certain relevance to the tasks above, but there are still some differences. The grammatical subject of a sentence in text summarization can only be regarded as a candidate for the subject of a sentence in this paper. In sentiment analysis, both the opinion holder and the opinion target can be the subject. The extraction of agents and patients in SRL is similar to subject recognition. The subject of a sentence can be either the agent or the patient. For example, the subject of the sentence "I'm mad at you"(我快要被你气死了) is the patient, while the subject of the sentence "I like you"(我喜欢你) is the agent. Therefore, all previous solutions can not handle the task of subject recognition issue directly.

## 3 Dataset Construction for Subject Recognition

As far as we know there is no open high-qulity subject annotation dataset in the field of chatbots. Due to the diversity of Chinese expression, irregularity of spoken Chinese, reference ambiguous or omitted of dialogue context and error propagation of speech recognition, it is difficult to form a common annotation guidelines for subject recognition. Based on a large number of real industrial chatbots interaction data annotated by human, this paper provides a relatively general subject annotation guidelines and constructs a dataset according to it. The annotation guidelines only for single sentences. All the data is derived from conversational corpus of users and chatbots in industrial products.

### 3.1 Subject Data Annotation Guidelines

**Categories of Subject Classification** The range of subjects discussed in this paper are all *limited*[4]. A subject can be one of the follows: personal pronouns,

[4] The term *limited* refers to specific reference and definite quantity, while *non-limited* means those generic reference and non-definite quantity. The subject discussed in

**Table 1.** General rules for subject annotation

| No. | Rule details | Example | Subject |
|---|---|---|---|
| 1 | A sentence startswith subject, has no object. Labeling grammatical subject. | *I am happy(我很开心).* | I |
| 2 | An SVO sentence, object is not a clause. Labeling grammatical subject. | *I like you(我喜欢你).* | I |
| 3 | An SVO sentence, object is a clause. Labeling grammatical subject of object clause. | *I think you are a fool(我觉得你很傻).* | You |
| 4 | A Verb-Object(VO) structure sentence, object is not a clause, the omitted subject can be determined, Labeling omitted subject. | *(I) was beaten by someone this morning(被人打了).* | I |
| 5 | A VO structure sentence, object is a clause, subject is the grammatical subject of the clause in most cases. | *(I) hope you are getting smarter(希望你越来越聪明)* | You |

name, properties, body parts of people, people belongings, some relationship of people and anthropomorphic animals. This paper defines five kinds of subject labels: "I", "You", "I & You", "Others" and "None".

**1) "I"** is the speaker of the sentence, including first person words (我/咱(I/my/me), 我们(our/us)), my properties, my body parts, my belongings and those sentences which omit "I" but is the sentence subject evidently.

**2) "You"** is the target of the speaker. It contains second person words (你/你们(you/ your)), the chatbot's name, your properties, your body parts, your belongings and those sentences which obviously omit "You" as subject.

**3) "I & You"** means the core topics or events of the sentence are done by both sides of the speaker and target. E.g. "I am as beautiful as you".

**4) "Others"** refers to someone else, neither the speaker nor the target of the speaker. Third person, person name (except chatbot's or user's name) or a particular person or group can regards as "Others".

**5) "None"** mainly contains: (1) The following components as subjects, such as objects, time words, place words, state words, abstract nouns and proper nouns, etc. (2) There is no subject in a sentence. (3) Sentence omits the subject and the omitted subject cannot be determined.

**General Annotation Rules** Based on a large number of manual annotation experiences, we abstract five general annotation rules which can be suitable for most scenarios. Rules are built on grammatical structures of sentences, as described in Table 1. It should be noted that not all sentences which satisfy the five structures above can be labeled. The key point is to first identify the core content of the sentence, and then label the core content's role as the subject.

**Default Supplement of Subject Omission** In the specific scenario of human-chatbot dialogue, we can supply a default subject to the sentences which omit the subject in some cases.

**1) Situations of adding "I" as the default subject**

this paper belongs to the limited, so the non-limited people are usually not regarded as subjects, such as the "men" in "All men must die".

- A VO structure sentence omits the subject and the object is a second person word. E.g. "Love you", "Thank you".
- A sentence omits the subject and describes the inner emotions and mental state of users. E.g., "I feel sad because I was hurt"(受伤了难受).

**2) Situations of adding "You" as the default subject**

- An imperative sentence which has a intention of chatbot function, such as setting alarms, playing music, etc. E.g. "Set a 10 o'clock alarm clock for me."
- A VO structure sentence omits the subject and the object is a first person. , such as "Give me a smile"(给我笑一个).

**3) Adding subject according to verb directionality**

Some verbs have obvious directionality, such as "把/ba", "被/bei", "给/gei", "帮/bang", "替/ti', "让/rang", "告诉/gao su", etc. In most cases, the subject of a sentence exists before these verbs. If the subject is omitted, it can be restored according to the verb direction and the personal pronouns of the object. E.g., "You amused me(被你逗乐了)" in Chinese omits the subject "I".

**Subject Choice in Sentence with Multiple Predicates** When a sentence contains juxtaposed components or is complex with clauses, there is usually more than one predicate. The following rules can be used to choose the most important one:(1) Prioritize the parts which contain chatbots function intents. (2) Give preference to the sub-sentence which describes the user's emotions.(3) Other juxtaposed situations such as causal relationship, and transition relationship, the subject of the highlighted part of the whole sentence can be chosen as the subject.

Another solution is to split each juxtaposed component into mutiple sentences and label subject for each sub-sentence. The rules above are all heuristic, readers should adjust in real practise.

### 3.2 Methods for Subject Dataset Construction

**Data Filtering** Due to the limitations of the accuracy in Speech Recognition, there exist a large number of mistakes in chatbot corpus that should be filtered. In general, the *Perplexity* calculated by language models can be a metric to evaluate the fluency of sentences. The formula of sentence *Perplexity* value is: $Perplexity(s) = \sqrt[N]{\frac{1}{P(w_1,w_2,\ldots,w_n)}} = \sqrt[N]{\frac{1}{\prod p(w_i)}}$, where $s$ represents a sentence, $N$ is the number of words in $s$, $p(w_i)$ means the probability of the $i$-th word. The smaller the *Perplexity* value is, the more fluent the sentence is. We use a traditional n-gram language model kenLM[5], and use large-scale news corpora and chatbot dialogue corpus which contain approximately 100 million sentences as the training data. In this paper, we use 500 *Perplexity* as the threshold to judge whether a sentence is fluent. The accuracy is 84.3% through manual sampling evaluation. After that, in order to ensure data diversity, we improve the Longest Common Subsequence (LCS) algorithm to filter similar data by ignoring punctuation, modal particles and other meaningless words in the sentence.

[5] https://github.com/kpu/kenlm

**Semi-Automatic Labeling** Annotation is an iterative process. We summarize and generalize some labeling rules based on a small part data of pre-labeling. Then, we tag data with different feature labels by these rules. After that, we put the data with similar labels together to annotate. It can greatly improve the efficiency and the consistency. We use the following features: number of subject words, the category of the subject word, whether the sentence starts with subject words, whether the sentence ends with subject words, the dependency relation "SBV"[6] in sentence, etc. We re-train the model on the annotate data and find out the underperformed sentences. Then, annotators reviewed them manually. Finally, a new model is trained on the checked data to predict unlabeled data.

### 3.3 Data Labeling Consistency Detection

Each data is annotated by two different annotators, the disagreement is judged by an auditor. We calculate a $Kappa$ value for the two annotators to get the label consistency. The formula is: $Kappa = \frac{p_o - p_e}{1 - p_e}$, where $p_o$ means the relative observed agreement among annotators, $p_e$ is the hypothetical probability of chance agreement. In general, the $Kappa > 0.75$ indicates a satisfying labeling consistency, while the $Kappa < 0.4$ indicates unsatisfied labeling consistency. Our dataset's $Kappa$ is 0.875. The result shows that the annotation guidelines in this paper is relatively uniform and less divergent. Labeling consistency and the quality of the data is high.

## 4 Subject Recognition of Chatbots

In this section, we first introduce a variety of baseline classification models based on deep neural networks. Then, according to the characteristics of the subject recognition task, the subject recognition models based on the pre-trained language models are constructed.

### 4.1 Subject Recognition Models Based on Deep Neural Network

In resent years, significant achievements have been made in text classification based on deep neural network models. In such models, words are usually embedded into fixed-length vectors. In this paper, a number of text classification models are introduced to verify the effectiveness of subject recognition based on the manual annotated data sets.

Kim [5] proposes a text classification model named TextCNN, based on features obtained from convolution kernels in a CNN network. Bi-LSTM model improves the traditional RNN network by incorporating context and temporal information Zhou [18] adds Attention mechanism to Bi-LSTM network to capture the core information in sentences by calculating temporal weights. Lai [6] combines the advantages of Bi-LSTM and CNN, and proposes RCNN (Recurrent Convolutional Neural Networks) model. The Adversarial LSTM model proposed

[6] A subject-verb relation in LTP. https://github.com/HIT-SCIR/ltp

by Miyato [8] generates a common training model for adversarial samples by randomly adding noise to word vectors to prevent over-fitting. Transformer [14] model abandons CNN and RNN, and uses only self-Attention mechanism, which greatly improves the computing speed while better captures the global information. FastText [3] is a classification model based on n-gram features. It has fast training speed and can maintain quite good performance.

### 4.2 Subject Recognition Based on ELMo

Contextual lexical information and word sequence have great influence on subject recognition. For example, "I like you" and "You like me" contain exactly the same vocabulary information, but the subjects are completely opposite because of the different word order. The word vectors obtained from traditional word vector models, such as Word2vec and GloVe, are static and unable to represent the differences between sentences. ELMo(Embedding from Language Models) [9] is a word vector model proposed by Allen NLP. It is essentially a bidirectional language model, in which words can be represented with context information. In real applications, word embeddings from ELMo can be dynamically adjusted in order to suit the subject recognition tasks.

To make full use of the context information and make better semantic representations, the ELMo word vectors are trained based on all data sets using Bi-LSTM with Attention. While fully mining the sequential information in text, the model can automatically capture the key information of words. Based on ELMo, three main subject recognition frameworks are constructed: 1) ELMo+Bi-LSTM; 2) ELMo+Bi-LSTM+Attention; 3) ELMo+TextCNN.

### 4.3 Subject Recognition Based on BERT

BERT (Bidirectional Encoder Representations from Transformers) [2] is a language representation model trained on large-scale text. It adopts Transformer as the main framework of the algorithm, which can better capture the bidirectional relationship in the corpus. The main feature of BERT is that it abandons the traditional RNN and CNN, and converts the distance between two words in any position to 1 through attention mechanism. It can effectively solve the long-term dependence problem in NLP tasks, and refreshes the highest record of 11 NLP tasks in GLUE benchmark at that time. Similar to ELMo, BERT can generate current word representation according to the semantics of the context words, which is very suitable for the subject recognition task in this paper.

## 5 Experiments

### 5.1 Dataset and Evaluation Metrics

We filtered duplicate conversational corpora with methods in Section 3.2. Totally 49,924 sentences are obtained and labeled with the guidelines in Section 3.1.

Readers can download the dataset on Github[7]. We divided the dataset into three parts: traning set(70%, 31,446), validation set(10%, 4,448) and test set(20%, 9,030). Additional 1,000 sentences as manual evaluation data set, which has no intersection with the above data sets.

We used the models listed in Section 4, and utilized Python3 as coding language. A pre-trained word embedding model is used in BERT, while in ELMo and fastText, word embeddings are trained on all data. For the others models, word embeddings were trained by Word2vec[8] with all data. We used the skip-gram architecture of Word2vec and the dimension is set to be 300.

We utilized Precision(P), Recall(R) and $F_1$ score as evaluation metrics. Particularly, due to the imbalance of the data with different labels, we used weighted P, R and $F_1$ score as final evaluation metrics. In order to verify the improvement of conversational dialogue in chatbots by subject identification, we did manual evaluation on three popular chatbots. Firstly, we got the replies through chatting with three chatbots on 1,000 test sentences. Secondly, we invited two language experts to annotate the subjects of these replies. The annotation methods are as follows:

- The subject of the reply is able to correspond the query. Labeling "correct".
- The subject of the reply is unable to correspond the query. Labeling "error".
- There is no subject in the reply but is still an appropriate answer. Labeling "notbad".
- The reply is not related to the query or a safe response. Labeling "unknown".

The higher ratio of "correct+notbad" indicates the better performance of chatbots, while the higher ratio of "error" reveals the worse performance on subject recognition. We assumed that if the subject recognition in the query is right, the subject of the reply will be correct. Therefore, we used subject recognition model to predict the subjects for the sentences, and then counted the number of "Error to Correct"(E2C), which denotes the number of sentences in which wrong subjects are transformed to be correct using subjects predicted by the model. We also defined a $Transformation$ *Ratio*(TR), as the ratio of the number of E2C, divided by the number of sentences with errors in subjects. This metric can reflect the improvement of conversational dialogue in chatbots by subject recognition task.

### 5.2 Experimental Results

**Models results** We carried out experiments on seven baseline models, three ELMo models and the BERT model, experimental results are shown in Table 2. The results show that TextCNN performed best (84.5% $F_1$) in models with Word2vec word embedding, and Transformer was slightly worse (81.9% $F_1$). The possible explanation is that the amount of data is not large enough. The results of other models were not much different. ELMo+Bi-LSTM was optimal among

[7] https://github.com/winnie0/ChineseSubjectRecognition
[8] https://radimrehurek.com/gensim/models/word2vec.html

**Table 2.** Experimental results of subject recognition

| Model | Val_P | Val_R | Val_$F_1$ | Test_P | Test_R | Test_$F_1$ |
|---|---|---|---|---|---|---|
| TextCNN | 85.2 | 84.6 | 84.9 | 84.8 | 84.4 | 84.5 |
| RCNN | 84.3 | 84.1 | 84.1 | 83.8 | 83.7 | 83.6 |
| Bi-LSTM | 84.6 | 83.6 | 84.0 | 83.8 | 82.9 | 83.3 |
| Bi-LSTM+Attention | 84.2 | 83.6 | 83.8 | 83.9 | 83.5 | 83.6 |
| AdversarialLSTM | 86.8 | 80.9 | 82.8 | 84.9 | 81.5 | 82.6 |
| Transformer | 82.3 | 81.5 | 81.9 | 82.3 | 81.7 | 81.9 |
| fastText | 83.0 | 83.4 | 83.0 | 83.0 | 83.4 | 83.1 |
| ELMo+Bi-LSTM | 85.0 | 85.4 | 85.1 | 85.6 | 85.4 | **85.5** |
| ELMo+Bi-LSTM+Attention | 84.7 | 85.1 | 84.8 | 85.3 | 85.5 | 85.3 |
| ELMo+TextCNN | 85.3 | 85.7 | 85.3 | 84.9 | 85.2 | 85.0 |
| BERT | 88.4 | 88.6 | 88.5 | 88.5 | 88.6 | **88.5** |

**Table 3.** Manual evaluation results of subject recognition

| chatbots | correct | error | notbad | unknown | correct R | error R | E2C | error dec R | TR |
|---|---|---|---|---|---|---|---|---|---|
| Bot A | 457 | 116 | 131 | 311 | 45.0% | 11.4% | 66 | 6.5% | 56.9% |
| Bot B | 418 | 88 | 222 | 287 | 41.2% | 8.7% | 63 | 6.2% | 71.6% |
| Bot C | 521 | 74 | 226 | 194 | 51.3% | 7.3% | 45 | 4.4% | 60.8% |

the three models which used the ELMo word embedding (85.5% $F_1$). It was useless to add Attention on ELMo+Bi-LSTM as shown by the results. We could see that all basic models increased $F_1$ by using the ELMo word embedding, which was improved by 2.2% on Bi-LSTM, 1.7% on Bi-LSTM+Attention and 0.5% on TextCNN. These results illustrated that the ELMo word embedding can capture the semantic information of sentences better and enhance the effect of subject recognition. However, the fine-tuning mode of the BERT model achieved excellent classification performance, which was superior to all other classification models, and reached 88.5% $F_1$ score.

**Manual evaluation** The manual evaluation results on 1,000 sentences of three popular chatbots are shown in Table 3. The model reduced the subject error rate by 6.5%, 6.2% and 4.4% in three chatbots respectively, and the transformation ratios were 56.9%, 71.6% and 60.8%, which were improved significantly.

## 6 Conclusion

In this paper, we present a newly built dataset for subject recognition in chatbots and propose a novel model for the task. Based on the data in real applications, we propose an annotation guidelines, and the manually annotated dataset contains 44,924 sentences. According to the characteristics of the subject recognition task, a number of classification models are proposed based on the pre-trained language models. Finally, we conduct experiments to verify the effectiveness of proposed models. In the future, we plan to expand the number of annotated data sets. Chatbots in real applications will also be improved based on the achievements to enhance the interaction experience with human.

## 7 Acknowledgements

The authors are supported by the National Natural Science Foundation of China (Grant No. 61603240). Corresponding author is Hao Shao. We also thank the anonymous reviewers for their insightful comments.

## References

1. Christensen, J., Soderland, S., Etzioni, O., et al.: An analysis of open information extraction based on semantic role labeling. In: K-CAP. pp. 113–120. ACM (2011)
2. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
3. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759 (2016)
4. Kim, S.M., Hovy, E.: Identifying opinion holders for question answering in opinion texts. In: AAAI-05 Workshop. pp. 1367–1373 (2005)
5. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)
6. Lai, S., Xu, L., Liu, K., Zhao, J.: Recurrent convolutional neural networks for text classification. In: AAAI (2015)
7. Li, F., Han, C., Huang, M., Zhu, X., Xia, Y.J., Zhang, S., Yu, H.: Structure-aware review mining and summarization. In: COLING. pp. 653–661 (2010)
8. Miyato, T., Dai, A.M., Goodfellow, I.: Adversarial training methods for semi-supervised text classification. arXiv preprint arXiv:1605.07725 (2016)
9. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. arXiv preprint arXiv:1802.05365 (2018)
10. Punyakanok, V., Roth, D., Yih, W.t.: The importance of syntactic parsing and inference in semantic role labeling. Computational Linguistics **34**(2), 257–287 (2008)
11. Qi, H., Yang, M., Meng, Y., Han, x., Zhao, T.: Skeleton parsing for specific domain chinese text (in Chinese). Journal of Chinese Information Processing **18**(1) (2004)
12. Qiu, G., Liu, B., Bu, J., Chen, C.: Opinion word expansion and target extraction through double propagation. Computational linguistics **37**(1), 9–27 (2011)
13. Rusu, D., Dali, L., Fortuna, B., Grobelnik, M., Mladenic, D.: Triplet extraction from sentences. In: IMSCI. pp. 8–12 (2007)
14. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in neural information processing systems. pp. 5998–6008 (2017)
15. Wang, R., Ju, J., Li, S., Zhou, G.: Feature engineering for crfs based opinion target extraction (in Chinese). Journal of Chinese Information Processing **26**(2) (2012)
16. Wiegand, M., Klakow, D.: Convolution kernels for opinion holder extraction. In: NAACL-HLT. pp. 795–803 (2010)
17. Zhou, H., Huang, M., Zhang, T., Zhu, X., Liu, B.: Emotional chatting machine: Emotional conversation generation with internal and external memory. In: AAAI (2018)
18. Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., Xu, B.: Attention-based bidirectional long short-term memory networks for relation classification. In: ACL (2016)