Emotion Analysis for the Upcoming Response in Open-Domain Human-Computer Conversation

Xiang Li and Ming Zhang(✉)

School of EECS, Peking University, Beijing, China
{lixiang.eecs,mzhang_cs}@pku.edu.cn

Abstract. Emotion analysis is one of the most active domains, hence attracts lots of attention of researchers in the natural language processing field. However, most of existed works are involved in classification tasks of the current sentence, lack of analysis of upcoming sentences. On the other hand, with the development of automatic human-computer dialogue systems, a response given by the computer side should become increasingly like human beings, for instance, the ability of expressing sentiment or emotion. The challenges lies in how to predict the emotion of a nonexistent sentence currently, which make this problem quite different from traditional sentiment or emotion analysis. In this paper, for the scenarios of open-domain conversation, we propose an architecture based on deep neural networks to predict the emotion before giving the response. In particular, we use a bidirectional recurrent neural network to get the embedding of the current utterance, and joint the representations of its retrieval results, to obtain the best emotion classification of the upcoming response. Experiments based on an annotation dataset demonstrate the effectiveness of our proposed approach better than traditional methods in terms of accuracy, precision, recall, and F-measure evaluation metrics. Then the following is some analysis of the results and future works.

Keywords: Emotion analysis · Deep learning · Neural networks
Open-domain · Human-computer conversation

1 Introduction

Emotion expression is one of the most important activities when human beings communicate with each other. For traditional face-to-face conversation, individuals often take delight in indicating their emotion by facial expression and tone. Recently with the prosperity of social media, emotion could be presented through many ways such as text including emotion words or emoji, and users are gradually accustomed to deliver their current emotion on web platforms like Facebook and Twitter, in order to make their messages more lively.

L. H. U and H. Xie (Eds.): APWeb-WAIM 2018, LNCS 11268, pp. 352–367, 2018.
https://doi.org/10.1007/978-3-030-01298-4_29

Table 1. Two examples of emotion in human-human conversation

Sentences	Emotion
Human1:好久没来看，帖子依旧在	none
(Long time no see, my post is still here)	
Human2:是呀，这感觉还不错吧	happiness
(Yes, it feels pretty good)	
Human1:这么明显的骗你，你都信！	anger
(So obvious a lie to you, unexpectedly you believe it!)	
Human2:他现在也已经不理我了，又消失了	sadness
(He ignores me now and disappears again)	

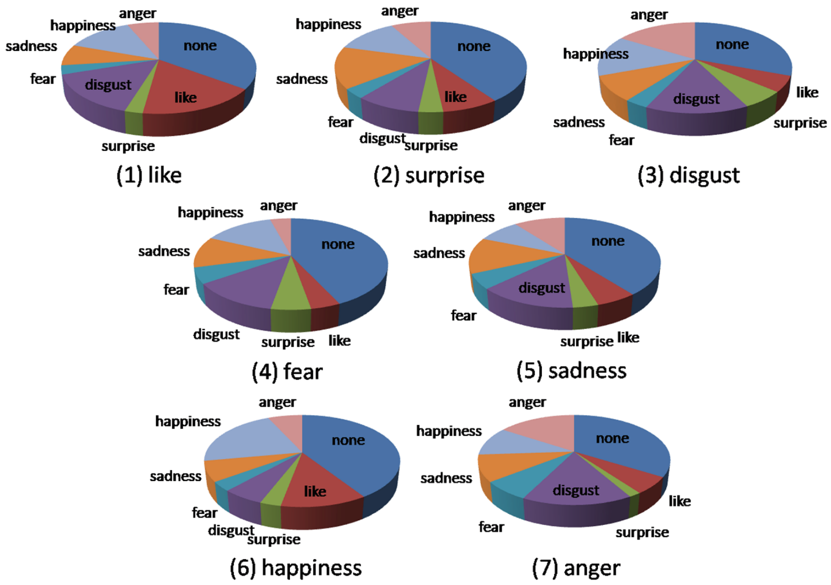
Hence, a large number of researchers in the natural language processing (NLP) field have paid attention to emotion analysis, which is usually a classification task, aiming to classify a span of text into one of several pre-defined emotion categories [1,17]. There are some different ways to define these categories according to affective science, and one of the most frequent way contains eight classes, i.e., like, surprise, disgust, fear, sadness, happiness, anger and none, especially for the emotion analysis of Chinese short text [1].

On the other hand, as one of the most challenging problems in artificial intelligence (AI), automatic human-computer dialogue systems have developed rapidly this several years. In order to make the computer side capable to communicate at the human level, it is necessary for the dialogue systems to express emotion proactively [2]. Thanks to the studies of controlled response generation, which means adding particular information into the generation process [3], it is possible to obtain an emotional response, if given the emotion category it should express [2]. Therefore, it becomes important to predict the correct emotion for the upcoming response before generation.

Under the scenarios of open-domain human-computer conversation, it is quite different from traditional sentiment or emotion analysis:

- The biggest challenge lies in that the prediction should be completed before generation, so that the result could be regarded as an extra input containing emotional information for the upcoming response, instead of classification on an existed sentence or document;
- Another point is that for open-domain scenarios, the content of conversation could refer to any area, and the utterances are usually with short length and casual expression, which make this task more challengeable [1];
- Last but not least, it is still lack of large-scale annotated dataset for emotion analysis, especially for the scenarios of open-domain human-computer conversation, which make supervised learning more difficult.

For traditional studies on sentiment or emotion analysis, most of them focus on classifying the current sentence. However, emotion of the upcoming response in conversation scenarios is quite different from the current utterance, so it is not suitable to simply propagate the same emotion category. As the examples showed in Table 1, under scenarios of human-human conversation, the second person may



(1) like (2) surprise (3) disgust

(4) fear (5) sadness

(6) happiness (7) anger

Fig. 1. For one-round conversation manually input by humans, we plot the proportion of emotion expressed in the first utterance, when given a particular emotion of the second person. The emotion consists of seven categories except none, i.e., like, surprise, disgust, fear, sadness, happiness, and anger. We aim to find out the relation of the emotion between the two utterances in one-round conversation. Generally speaking, most of the emotions expressed in the second utterances are different from its context, which make the problem more complicated.

start to present his emotion (as the first example, from none to happiness), or there probably is an emotional transformation with the conversation proceeding (as the second example, from anger to sadness). Moreover, in order to find out this difference quantitatively, we also investigate over one thousand pieces of one-round human-human conversation, which contain emotional expression in the sentence given by the second person according to human annotation. As shown in Fig. 1, our observation is that, only a small percentage of sentences maintain the same emotion category with their previous utterances. Therefore, it is necessary to introduce extra information to predict the upcoming response in conversation scenarios.

Thanks to the maturity of information retrieval (IR), we could collect the requisite information from retrieval-based methods. With the development of social media, accumulate numbers of individuals send posts and reply other people on these platforms like Twitter[1] and Baidu Tieba[2]. These kinds of resource could be regarded as human-human conversation, so that it is possible to construct data collections with large scales for human-computer dialogue systems. Since

[1] http://www.twitter.com.

[2] https://tieba.baidu.com.

conversation tasks could be transformed to information retrieval scenarios, one kind of methods to get the upcoming response is selecting an existing sentence from a large scale of corpus as the best reply to the input post [4,5]. The advantages of retrieval-based conversation systems are that the computer side could behave almost like a human and the replies usually have high correlation with the input posts.

Thus, in this paper, we take advantage of the information in retrieval results, to predict the emotion of the upcoming response. Our motivation is that retrieval results could be regarded as candidate responses, which indicate the directions of the final upcoming response, so they probably contain key information of the upcoming emotion that is difficult to be found out only according to the current utterance. Therefore, it is natural to combine the current utterance and its retrieval results. To be specific, we propose a novel architecture based on deep neural networks to deal with this combination and learn emotion classification. Based on the layer of word embeddings, we get the embedding of the current utterance using a bidirectional recurrent neural network (RNN) improved by Long Short-Term Memory (LSTM), and then concatenate it with the sentence embeddings of retrieval results, which are also obtained by a bidirectional LSTM model. Finally, we use full connection layers and a softmax function to complete the emotion classification. Comparing to the traditional methods of emotion analysis, our approach could integrate information from retrieval results and better predict the emotion category.

To sum up, the main contributions of this paper are as follows.

- To the best of our knowledge, we are the first to address the problem of emotion analysis for the upcoming response under open-domain human-computer conversation scenarios, which is quite different from traditional emotion classification.
- We propose a novel classification model using deep neural networks to solve this problem, combining the current utterance and its retrieval results into a hybrid framework, for the purpose of obtaining more information of the upcoming response.
- Empirical experiments demonstrate the effectiveness of our approach, with a better performance competing to traditional methods in terms of different evaluation metrics.

2 Related Work

2.1 Conversation Systems

With the increasing effect to people's everyday life, open-domain human-computer conversation systems have attracted much attention of industrial and academic communities. Many state-of-the-practice systems are developed and have amounts of users, including Siri of Apple, Xiaobing of Microsoft, and Dumi of Baidu. There is also a rising trend for academic studies to tackle the correlation and flexibility challenges, mainly with retrieval-based or generative methods.

Information retrieval is a classical topic which has developed for decade years. The traditional task is getting a ranked list of documents related to the given query, with a process consisting of pre-filter and re-ranking [27]. After retrieving candidates from a large scale of documents, features such as latent structures [6] are calculated to execute a ranking model, like learn-to-rank [7] or semantic matching [8]. When considering scenarios of human-computer conversation as selecting the most appropriate response for the current utterance, it is natural to adapt information retrieval technology for both traditional vertical dialogue systems and open-domain conversation. Leuski and Traum developed a virtual human dialogue system named NPCEditor, using information retrieval techniques [9]. Higashinaka et al. proposed a method to filter noise for candidate sentences in dialogue systems, with techniques of syntactic filtering and content-based retrieval [10]. Retrieval-based approaches could search an enough intelligent response while combining with high-quality and large-scale of conversation data resources [5,11].

In addition to information retrieval, another kind of methods is based on generation. Compare with retrieval-based methods that use instances already existing, generation-based ones are regarded more flexible. Besides filling templates [12] and paraphrase generating [13], statistical machine translation and neural networks are two main kinds of technologies to generate the best response. Based on techniques of Statistical Machine Translation, Langner et al. generate replies by translating the internal dialogue state [14], while Ritter et al. propose a data-driven approach to generate the reply for posts in Twitter [15]. Recently, with neural networks and deep learning being demonstrated useful for many natural language processing tasks, recurrent neural networks are used in conversation systems increasingly, for example, Shang et al. formalize the generation of reply as a decoding process and used recurrent neural networks for both encoding and decoding [16]. The encoder-decoder framework based on neural networks are also adapted to other complex scenarios, such as context-sensitive generation in conversation systems [18].

Since the technology of information retrieval could be used to obtain the response under conversation scenarios and demonstrated effective, it is natural for us to believe that the retrieval results could indicate possible directions of the real upcoming response and exploit key information in them.

2.2 Sentiment Analysis

Traditional sentiment or emotion analysis is a significant research task which has attracted many researchers in the domain of natural language processing. Research work on sentiment analysis often focuses on classifying the polarities of positive and negative, or extends to the third polarity of neutral, or sometimes adds fine-grained classes like a spectrum such as very positive and very negative. Emotion analysis is a kind of classification task aiming to distinguish several predefined emotion categories, such as happy, sad, and so on, while sometimes the emotion classification could be multi-label.

With the development of natural language processing, many theories and technologies have been used to deal with traditional sentiment or emotion analysis. Since some words could provide clear clues of sentiment or emotion classification, an effective series of approaches is lexicon-based models, which rely heavily on dictionaries [19,20]. Another kind of methods with high performance is feature-based models using traditional classifiers, which is called distant supervision by leveraging sentences or documents with human annotation. Support vector machine and conditional random fields based machine learning models are used to implement the emotion classification of web blog corpora [21]. Wen and Wan mined class sequential rules to improve performance of the support vector machine for emotion classification in microblog texts [1]. Other theories like statistical machine translation [22], graph-based approach [24] and topic model [23,25,39] are also used to deal with this kind of tasks.

Recent years, with the development neural networks, research work appears continuously using deep learning approaches to improve the performance of tasks in the natural language processing field, including sentiment or emotion analysis. Since semantic features and latent information could be represented by its embedding, one series of methods is to add specific information indicating sentiment or emotion categories into the word embeddings while training by neural networks [26,28,29]. Proposing novel structures of neural networks is another kind of approaches with high performance [30–33], which means adapting the theory of deep learning to these tasks. Furthermore, context of human interaction are considered to improve sentiment or emotion analysis under some specific scenarios, especially on social networks [34–36].

Under human-computer conversation scenarios, there are also some research works about sentiment or emotion topics. A neural learning approach is proposed to estimate the sentiment of the upcoming response in dialogue systems, while it only distinguish the sentiment polarities of positive and negative, or adding neutral as extension, instead of several categories of emotion, and it only consider about the information in the conversation process, without extra information such as retrieval results [37]. Another series of works is emotional conversation generation, as one kind of controlled response generation, aiming to gift the computer side ability of expressing emotion [2,38,40]. However, this kind of works regard the emotion category as a given input, instead of calculating it.

To the best of our knowledge, under scenarios of open-domain human-computer conversation, it still lacks works of emotion analysis for the upcoming response, and the prediction result could be regarded as the input of emotional conversation generation. Therefore, we propose a novel classification model utilizing retrieval results to solve this problem.

3 Approach

3.1 Task Definition

Given the current utterance $X = \{x_1, x_2, ..., x_T\}$, our aim is to train a classification model which could predict the emotion possibility $P(y|X), y \in Y$ for

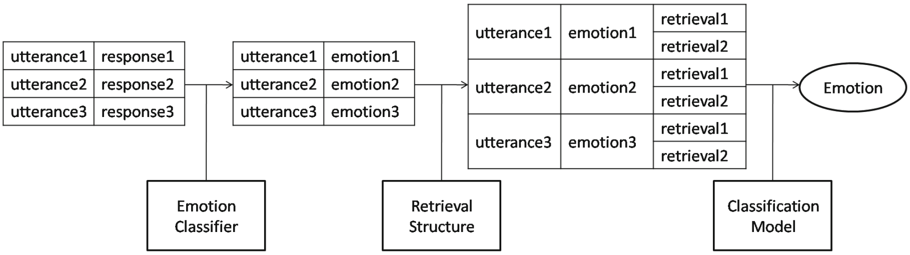


Fig. 2. The whole process to classify the emotion category of the upcoming response.

the upcoming response of the utterance X, under open-domain human-computer conversation scenarios. Y is the pre-defined emotion set mentioned before, e.g., $Y = \{like, surprise, disgust, fear, sadness, happiness, anger, none\}$.

3.2 Overview

The whole process of our proposed approach is shown in Fig. 2, which generally consists of three parts.

- Due to the lack of large-scale annotated dataset for emotion analysis, especially for the scenarios of open-domain human-computer conversation, the first part is a traditional emotion classifier to label the emotion of the response for each data item. The original training data consist of pairs of (utterance, response), and after the classification, they will become pairs like (utterance, emotion), since actually, we only need the emotion label of the response instead of its content. Here we use a bidirectional LSTM model as the classifier, which has better performance than traditional methods [2].
- Then, the second part is a retrieval structure, to get the retrieval results of the current utterance. Hence, after the retrieval process, each data item will be extended to tuples like (utterance, emotion, retrieval results), as the final input of our proposed classification model, noting that the emotion label is needed only in the training process.
- For the final part, it is our proposed classification model based on deep neural networks, to predict the emotion of the upcoming response. The model calculate the sentence embeddings of the current utterance and its retrieval results respectively, and deliver their combination to a softmax layer.

3.3 Bidirectional LSTM

Since natural language sentences could be regarded as sequences, it is natural to use recurrent neural networks to model sentences and get their embeddings. For each hidden layer, the inputs are the current word embedding as well as the last hidden layer, until the end of the sentence, and the final hidden layer is regarded as the embedding of the whole sentence, which could represent all the sequential information. In practice, due to the increasing sparsity with the propagation going on, the Long Short-Term Memory (LSTM) [41] is often used to improve its performance.One LSTM unit could be regarded as a hidden state in the RNN structure, which could remember the information of words far away from the current word in the sentence. The specific calculation is given by

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2}$$

$$C'_t = tanh(W_C[h_{t-1}, x_t] + b_C) \tag{3}$$

$$C_t = f_t * C_{t-1} + i_t * C'_t \tag{4}$$

$$o_t = \sigma(W_o[h_{h-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * tanh(C_t) \tag{6}$$

where W's are weights and b's are bias terms. x_t is the word embedding; h_t is the hidden state at time step t; and the signal "*" denotes element-wise product of two vectors.

Bidirectional LSTM is an important variance of the RNN structure, which has been demonstrated with high performance under lots of scenarios [42].For each time step t, there are two hidden states, both connected with the input layer and the output layer. Thus, it could be regarded as two RNN chains that one propagates from the beginning to the end of the sentence, and another is from the end to the beginning, sharing the same input and deciding the output together. The advantage of this structure is not only to utilize the past information of the current word, but also the future information after it.

Therefore, bidirectional LSTM has high performance on traditional emotion classification task and we directly use it to label the response as mentioned before. Besides, it also is an important part of our proposed classification model to predict the emotion of the upcoming response, which will be introduced later. Instead of giving output at each time step, we only concatenate the final hidden states of the two RNN chains as the sentence embedding in this paper, because our classification scenarios aim at whole sentences.

3.4 Retrieval Structure

To get the retrieval results of the current utterance, we use a retrieval framework based on textual similarity to search out k candidate responses just like retrieval-based dialogue systems, which are semantically related to the utterance and could indicate the directions of the final upcoming response. In this part, we describe in detail the retrieval-based section, which adapts typical frameworks in the domains of search engine or advertisement selection, as shown in Fig. 3, working in a two-step retrieval-and-ranking strategy.

To start, the only input is the current utterance, and we use it to retrieve up to ten hundred candidate replies, from a large scale of conversation campus. The campus consists of post-reply pairs like one-round conversation, so the candidates

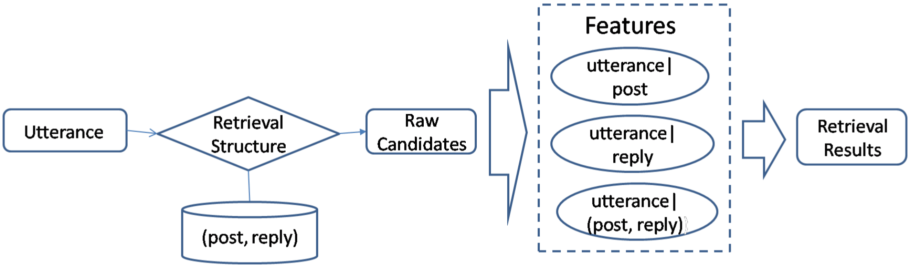


Fig. 3. The retrieval structure using a two-step strategy, the same with typical frameworks in the domain of information retrieval.

picked up should have at least one post textually similar to the utterance. This step is accomplished based on standard keyword retrieval structures, similar to the Lucene[3] and Solr[4] system.

After the coarse-grained retrieval, we have a set of raw candidate replies, and next we need to rerank these candidates in a fine-grained fashion, using much richer information from different perspective. Semantic meanings of a post and its reply are both important, so features should be exacted respectively focusing on three aspects, which means between the utterance and the post, the reply, and the post-reply pair, including textual similarity, measures of word embeddings, and so on. Finally, we get a ranking model and then use it to get top k replies, as the retrieval results of the current utterance.

3.5 Classification Model

Having the retrieval results of the current utterance, naturally the most important part is our proposed classification model to predict the emotion category of the upcoming response. As shown in Fig. 4, the whole structure model the current utterance and its retrieval results respectively and combine them together to make the final classification, based on deep neural networks.

For the left part in Fig. 4, we aim to obtain sentence embedding of the current utterance. A word embedding layer is at the bottom to represent input words in the utterance, and then put into a bidirectional LSTM structure introduced before, given by

$$s_t^{u+} = LSTM(s_{t-1}^{u+}, e(x_t)) \quad and \quad s_t^{u-} = LSTM(s_{t+1}^{u-}, e(x_t)) \tag{7}$$

where x_t is the input word and $e(x_t)$ indicates its embedding. "u" in the superscript means modeling the utterance; "+" means the direction from the beginning to the end; yet "−" is the opposite. Then, we concatenate the two final hidden states of the two directions and after an extra full-connection layer, we

[3] http://lucene.apache.org.
[4] http://lucene.apache.org/solr.

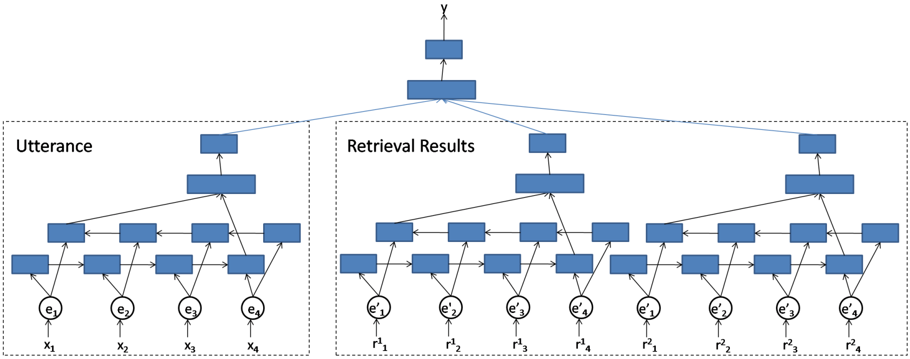


Fig. 4. The structure of our proposed model to predict the emotion category of the upcoming response. Note that there are k retrieval replies on the right part, and we only figure out two for simplicity.

complete the modeling process of the current utterance and get its final sentence embedding s^u, e.g.

$$s^u = \sigma(W_u[s_T^{u+}; s_1^{u-}] + b_u) \tag{8}$$

where W_u is the weight matrix, b_u is the bias term, and ";" is the concatenation operation of two vectors.

For the modeling process of the retrieval results, as the right part in Fig. 4 shown, the calculation is similar to that for the utterance, which means

$$s_t^{i+} = LSTM(s_{t-1}^{i+}, e'(r_t^i)) \quad and \quad s_t^{i-} = LSTM(s_{t+1}^{i-}, e'(r_t^i)) \tag{9}$$

$$s^i = \sigma(W_r[s_T^{i+}; s_1^{u-}] + b_r) \tag{10}$$

where r's are words in retrieval results and "i" in the superscript indicates the i-th retrieval result, with a range of 1 to k.

Then, we combine the two parts through concatenating sentence embeddings of the current utterance and all k retrieval results, and after another full-connection layer, we put it into the final softmax layer to make the emotion classification, e.g.

$$P(y|X) = softmax(W_2\sigma(W_1[s^u; s^1; ...; s^k] + b_1) + b_2) \tag{11}$$

4 Experiments

4.1 Datasets and Setups

For the whole process of our emotion analysis, there are 4 datasets as follows.

- Conversation Dataset. To construct the index base for our retrieval section, we collected massive resources from Chinese forums, microblog websites, and community QA platforms such as Baidu Zhidao, Baidu Tieba, Douban forum, Sina Weibo[5], and totally extracted nearly 10 million post-reply pairs.

- NLPCC Datasets. These two datasets were used in challenging tasks of emotion classification in NLPCC2013 and NLPCC2014[6]. We use them to train a traditional emotion classification model with bidirectional LSTM to label the replies of our training data.
- Training Dataset. Over 1.3 million post-reply pairs collected the same way as the conversation dataset, with emotion labels of the replies given by the bidirectional LSTM classifier.
- Test Dataset. Also post-reply pairs with emotion labels of the replies. Note that the emotion labels here are annotated by humans as the "ground truth", different from those of training dataset. Each item is annotated by 3 individuals in an independent and blind fashion, and finally we get an annotated dataset consisting of 1996 pairs, with eight categories mentioned before, and its kappa score $\kappa = 0.427$, showing moderate inter-rater agreement.

Metrics. We use several evaluation metrics to demonstrate the effectiveness of our proposed model. The first one is accuracy, which could indicate the correctness directly. Note that the type of none is less meaningful, yet occupies a high proportion. So more important series of metrics is precision, recall and F-measure, which is usually used for emotion analysis [1].

Training Settings. For the training process, we use cross-entropy objective as the loss function and all dimensions of embedding vectors are 128. Different retrieval results should share the same variate to be trained including word embeddings, parameters of bidirectional LSTM and the full connection layer in our model design, yet different from those for modeling the current utterance. However, for the consideration of running time, finally we make k equal to 1.

4.2 Baseline Algorithms

To demonstrate the effectiveness of our proposed model, we include the following methods as baselines. Besides traditional approaches, we also use some basic neural network structures to model the current utterance and make classification. For fairness, we conduct the same data cleaning and layer dimensions in neural networks for all algorithms. Specifically, we filter out utterances containing words with very low frequency, and also utterances containing over 50 words. Basic data pre-processing is also done, including word segmentation and so on.

SVM. SVM is one kind of traditional classification model to construct hyperplanes according to pre-defined features. Here, it is primarily based on filtered word features, and also some secondary features such as the emotion category of the current utterance given by the traditional emotion classifier. Besides, we use its linear version due to the large scale of our training data.

[5] http://zhidao.baidu.com, http://tieba.baidu.com, http://douban.com, http://weibo.com.

[6] http://tcci.ccf.org.cn/conference/2013|2014/.

Table 2. Performance of the emotion classification for the upcoming response

Method	Accuracy	Macro average			Micro average		
		Precision	Recall	F-measure	Precision	Recall	F-measure
SVM	0.4118	0.1669	0.0283	0.0484	0.2222	0.0335	0.0581
CNN	**0.4284**	0.1571	0.0021	0.0041	0.1667	0.0026	0.0052
LSTM	0.4238	0.2413	0.0165	0.0309	0.2639	0.0167	0.0314
Bi-LSTM	0.4279	0.3231	0.0183	0.0346	0.2658	0.0185	0.0346
Our method	0.4259	**0.3892**	**0.0786**	**0.1308**	**0.4026**	**0.0819**	**0.1361**

CNN. Convolutional neural networks is a kind of structure good at extracting local features. Instead of the way of full-connection, a neuron in the convolutional layer could only have particular numbers of connections from the last layer. After the convolutional operation, there is often a pooling layer to integrate the information. Specifically, the convolutional layer has 128 filters, with a window size of 3, and here we use max pooling.

LSTM. A recurrent neural network structure improved by Long Short-Term Memory units, which is introduced before.

Bi-LSTM. The variance of LSTM with two recurrent directions, which is also introduced before.

4.3 Performance

In this section, we show the performance of our proposed model against other baselines, and report the performance of emotion classification in all the mentioned evaluation metrics in Table 2. Since retrieval results could give key information for the real upcoming response in some degree, it is obvious that our method utilizing retrieval results perform better than those only with the current utterance, whatever SVM or basic neural network structures, with a comparable accuracy and higher precisions, recalls, and F-measures. However, sometimes the retrieval results are general replies without any emotional information, such as “I think so”, so the recall is always lower than the precision, which have large room for improvement.

4.4 Analysis

To analyze the performance of our proposed model specifically, we also investigate the F-measure for each emotion category. Some emotion categories have higher F-measures than the macro-average F-measure, while others not, especially for the “fear” emotion. The reason is that it occupies only a low proportion in practical conversation, so that could not be adequately trained. Thus, our observation is that emotion with higher appearance tends better performance, such as the “like” emotion, with a higher percentage than the other six emotion categories.

Table 3. A case study of the emotion classification for the upcoming response

Utterance	我在等学校装空调 (I am waiting for the air conditioner installed by the school)	None
Response	什么学校哇，这么好(What school? So nice)	Like
Retrieval	好牛(Awesome)	Like

Table 3 shows a case study, in which the current utterance is just a statement without any clear emotion, but the upcoming response comes up with "like" emotion with the expression of "so nice" to the "school". There is no clue to deduce this emotion only from the content of the utterance, yet the retrieval result could indicate that its response will probably be in the "like" emotion category.

5 Conclusion

In this paper, under open-domain human-computer conversation scenarios, in order to solve the problem of emotion analysis for the upcoming response, we propose an approach of jointing representations of the current utterance and its retrieval results using deep neural networks, and deeply analyze its performance through experiments. Empirical results demonstrate our approach better than traditional methods in terms of different metrics. For the future work, one direction is to propose more progressive models to consider contextual information in the conversation process, and another may be a global model to joint the emotion analysis and the controlled response generation, in stead of giving the emotion before the generation process.

Acknowledgements. National Natural Science Foundation of China NSFC Grant (NSFC Grant Nos.61772039, 91646202 and 61472006).

References

1. Wen, S., Wan, X.: Emotion classification in microblog texts using class sequential rules. In: AAAI Conference on Artificial Intelligence, pp. 187–193 (2014)
2. Zhou, H., Huang, M., Zhang, T., Zhu, X., Liu, B.: Emotional chatting machine: emotional conversation generation with internal and external memory. arXiv preprint arXiv:1704.01074 (2017)
3. Mou, L., Song, Y., Yan, R., Li, G., Zhang, L., Jin, Z.: Sequence to backward and forward sequences: a content-introducing approach to generative short-text conversation. In: International Conference on Computational Linguistics, pp. 3349–3358 (2016)
4. Yan, R., Song, Y., Wu, H.: Learning to respond with deep neural networks for retrieval based human-computer conversation system. In: Proceedings of the SIGIR, pp. 55–64 (2016)

5. Ji, Z., Lu, Z., Li, H.: An information retrieval approach to short text conversation. arXiv preprint arXiv:1408.6988 (2014)
6. Wu, W., Lu, Z., Li, H.: Learning bilinear model for matching queries and documents. J. Mach. Learn. Res. **14**(1), 2519–2548 (2014)
7. Liu, T.Y.: Learning to rank for information retrieval. Found. Trends Inf. Retr. **3**(3), 225–331 (2009)
8. Giunchiglia, F., Yatskevich, M., Shvaiko, P.: Semantic matching: algorithms and implementation. J. Data Semant. **IX**, 1–38 (2007). Springer, Berlin, Heidelberg
9. Leuski, A., Traum, D.: Npceditor: creating virtual human dialogue using information retrieval techniques. AI Mag. **32**(2), 42–56 (2011)
10. Higashinaka, R., et al.: Syntactic filtering and content-based retrieval of twitter sentences for the generation of system utterances in dialogue systems. In: Rudnicky, A., Raux, A., Lane, I., Misu, T. (eds.) Situated Dialog in Speech-Based Human-Computer Interaction. SCT, pp. 15–26. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-21834-2_2
11. Wang, H., Lu, Z., Li, H., Chen, E.: A dataset for research on short-text conversations. In: Proceedings of the conference on human language technology and empirical methods in natural language processing, pp. 935–945 (2013)
12. Sugiyama, H., Meguro, T., Higashinaka, R., Minami, Y.: Open-domain utterance generation for conversational dialogue systems using web-scale dependency structures. In: Proceedings of the SIGDIAL, pp. 334–338 (2013)
13. Mairesse, F., Young, S.: Stochastic language generation in dialogue using factored language models. Comput. Linguist. **40**(4), 763–799 (2014)
14. Langner, B., Vogel, S., Black, A. W.: Evaluating a dialog language generation system: comparing the mountain system to other NLG approaches. In: Annual Conference of the International Speech Communication Association (INTERSPEECH), pp. 1109–1112 (2010)
15. Ritter, A., Cherry, C., Dolan, W. B.: Data-driven response generation in social media. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 583–593 (2011)
16. Shang, L., Lu, Z., Li, H.: Neural responding machine for short-text conversation. In: Annual Meeting of the Association for Computational Linguistics, pp. 1577–1586 (2015)
17. Li, X., Yan, R., Zhang, M.: Joint emoji classification and embedding learning. In: Chen, L., Jensen, C.S., Shahabi, C., Yang, X., Lian, X. (eds.) APWeb-WAIM 2017. LNCS, vol. 10367, pp. 48–63. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63564-4_4
18. Sordoni, A., et al.: A neural network approach to context-sensitive generation of conversational responses. In: The Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 196–205 (2015)
19. Turney, P. D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Annual Meeting of the Association for Computational Linguistics, pp. 417–424 (2002)
20. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Comput. Linguist. **37**(2), 267–307 (2011)
21. Yang, C., Lin, K. H. Y., Chen, H. H.: Emotion classification using web blog corpora. In: IEEE/WIC/ACM International Conference on Web Intelligence, pp. 275–278 (2007)
22. Lambert, P.: Aspect-level cross-lingual sentiment classification with constrained SMT. In: Annual Meeting of the Association for Computational Linguistics, pp. 781–787 (2015)

23. Hai, Z., Cong, G., Chang, K., Liu, W., Cheng, P.: Coarse-to-fine review selection via supervised joint aspect and sentiment model. In: Proceedings of the SIGIR, pp. 617–626 (2014)
24. Wang, X., Wei, F., Liu, X., Zhou, M., Zhang, M.: Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. In: Proceedings of the CIKM, pp. 1031–1040 (2011)
25. Yang, M., Peng, B., Chen, Z., Zhu, D., Chow, K.P.: A topic model for building fine-grained domain-specific emotion lexicon. In: Annual Meeting of the Association for Computational Linguistics, pp. 421–426 (2014)
26. Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., Qin, B.: Learning sentiment-specific word embedding for twitter sentiment classification. In: Annual Meeting of the Association for Computational Linguistics, pp. 1555–1565 (2014)
27. Li, X., Mou, L., Yan, R., Zhang, M.: Stalematebreaker: a proactive content-introducing approach to automatic human-computer conversation. In: International Joint Conference on Artificial Intelligence, pp. 2845–2851 (2016)
28. Zhou, H., Chen, L., Shi, F., Huang, D.: Learning bilingual sentiment word embeddings for cross-language sentiment classification. In: Annual Meeting of the Association for Computational Linguistics, pp. 430–440 (2015)
29. Ren, Y., Zhang, Y., Zhang, M., Ji, D.: Improving twitter sentiment classification using topic-enriched multi-prototype word embeddings. In: AAAI Conference on Artificial Intelligence, pp. 3038–3044 (2016)
30. Tang, D., Wei, F., Qin, B., Zhou, M., Liu, T.: Building large-scale twitter-specific sentiment lexicon: a representation learning approach. In: International Conference on Computational Linguistics, pp. 172–182 (2014)
31. Dong, L., Wei, F., Zhou, M., Xu, K.: Adaptive multi-compositionality for recursive neural models with applications to sentiment analysis. In: AAAI Conference on Artificial Intelligence, pp. 1537–1543 (2014)
32. Zhang, M., Zhang, Y., Vo, D. T.: Gated neural networks for targeted sentiment analysis. In: AAAI Conference on Artificial Intelligence, pp. 3087–3093 (2016)
33. Zhao, Z., Liu, T., Hou, X., Li, B., Du, X.: Distributed text representation with weighting scheme guidance for sentiment analysis. In: Li, F., Shim, K., Zheng, K., Liu, G. (eds.) APWeb 2016. LNCS, vol. 9931, pp. 41–52. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45814-4_4
34. Vanzo, A., Croce, D., Basili, R.: A context-based model for sentiment analysis in twitter. In: International Conference on Computational Linguistics, pp. 2345–2354 (2014)
35. Ren, Y., Zhang, Y., Zhang, M., Ji, D.: Context-sensitive twitter sentiment classification using neural network. In: AAAI Conference on Artificial Intelligence, pp. 215–221 (2016)
36. Li, S., Huang, L., Wang, R., Zhou, G.: Sentence-level emotion classification with label and context dependence. In: Annual Meeting of the Association for Computational Linguistics, pp. 1045–1053 (2015)
37. Bothe, C., Magg, S., Weber, C., Wermter, S.: Dialogue-based neural learning to estimate the sentiment of a next upcoming utterance. In: Lintas, A., Rovetta, S., Verschure, P.F.M.J., Villa, A.E.P. (eds.) ICANN 2017. LNCS, vol. 10614, pp. 477–485. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68612-7_54
38. Zhang, R., Wang, Z., Mai, D.: Building emotional conversation systems using multi-task Seq2Seq learning. In: Huang, X., Jiang, J., Zhao, D., Feng, Y., Hong, Y. (eds.) NLPCC 2017. LNCS (LNAI), vol. 10619, pp. 612–621. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73618-1_51

39. Hai, Z., Cong, G., Chang, K., Cheng, P., Miao, C.: Analyzing sentiments in one go: a supervised joint topic modeling approach. IEEE Trans. Knowl. Data Eng. **29**(6), 1172–1185 (2017)
40. Yuan, J., Zhao, H., Zhao, Y., Cong, D., Qin, B., Liu, T.: Babbling - the HIT-SCIR system for emotional conversation generation. In: Huang, X., Jiang, J., Zhao, D., Feng, Y., Hong, Y. (eds.) NLPCC 2017. LNCS (LNAI), vol. 10619, pp. 632–641. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-73618-1_53
41. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
42. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw. **18**(5–6), 602–610 (2005)